CytoSure™ Interpret software

User guide

version 11

This protocol booklet and its contents are © Oxford Gene Technology (Operations) Limited 2008. All rights reserved. Reproduction of all or any substantial part of its contents in any form is prohibited except that individual users may print or save portions of the protocol for their own personal use. This licence does not permit users to incorporate the material or any substantial part of it in any other work or publication, whether in hard copy or electronic or any other form. In particular (but without limitation) no substantial part of the protocol booklet may be distributed or copied for any commercial purpose.

Contents

1. Computer requirements, software installation and updates……...7

2. Running CytoSure analysis software……...7

3. Loading data files……..10

4. Viewing CytoSure data……14

4.1 Navigation……..15
4.2 Annotation……..20
4.2.1 Adding custom annotation tracks
4.2.2. Filtering the annotation
4.3 Searching for a **gene** or **probe** or **syndrome**……..34
4.4 Data display (the radio buttons)……..35
4.5 Filtration of the data……..35
4.6 Normalisation and smoothing……..37
4.7 Multiple datasets……...39
4.8 Customising the display……...39
4.9 Table view tab……44
4.10 Aberration tab and ideogram view……..45
4.11 Aneuploidy testing……47

5. Adding aberrations to generate an aberration list, annotating, saving and exporting……..48

5.1 Automated aberration detection……..48
5.2 Batching automated aberration detection……..53
5.3 Manual aberration detection……...55
5.4 Link to SUSPECTS……...57
5.5 Editing and annotating the aberration list and saving……..60
5.5.1. Customising the classification terms

5.6 Exporting the aberration results for printing……..70
5.6.1. Customising the report……..……....71
5.7 Exporting the aberration results in Decipher format……75
5.8 P-value……75

6. Loop experimental design……...76

6.1 Import the 3 arrays and run the CBS……77

6.2 Checking the aberrations manually……56
6.3 Running the loop analysis……...57
6.4 Manually examine replicates……...60
6.5 Rerunning the loop analysis……62
6.6 Saving loop aberrations to the database……...62
6.7 Alternative method: The combination methods……63
6.8 Saving the loop aberrations to the database – combination method……65

7. Database Management……66

7.1 Positioning the database file in a location of the user's choice……..66
7.2 Viewing the data within the database……68
7.3 Deleting aberrations in the internal database……...72
7.4 Editing aberrations in the internal database……...74
7.5 Sorting or filtration of data in the internal database……..76
7.6 Displaying QC metrics……..77
7.7 Backing up the database file……...78
7.8 Exporting the data from the database file……79

8. Population analysis……..79

8.1 Grouping data and plotting aberration frequency……79
8.2 Combining groups……83
8.3 Deleting groups……85
8.4 Exporting aberration frequency results and statistics……..86

9. Setting up a protocol in Workflow mode……87

10. Use of the software in Workflow mode

11. Other features……87

12. Contact details……88

13. Legal information……..88

Introduction

The CytoSure Interpret software is a specialist Cytogenetic software package designed for the analysis of array CGH by Cytogenetists.

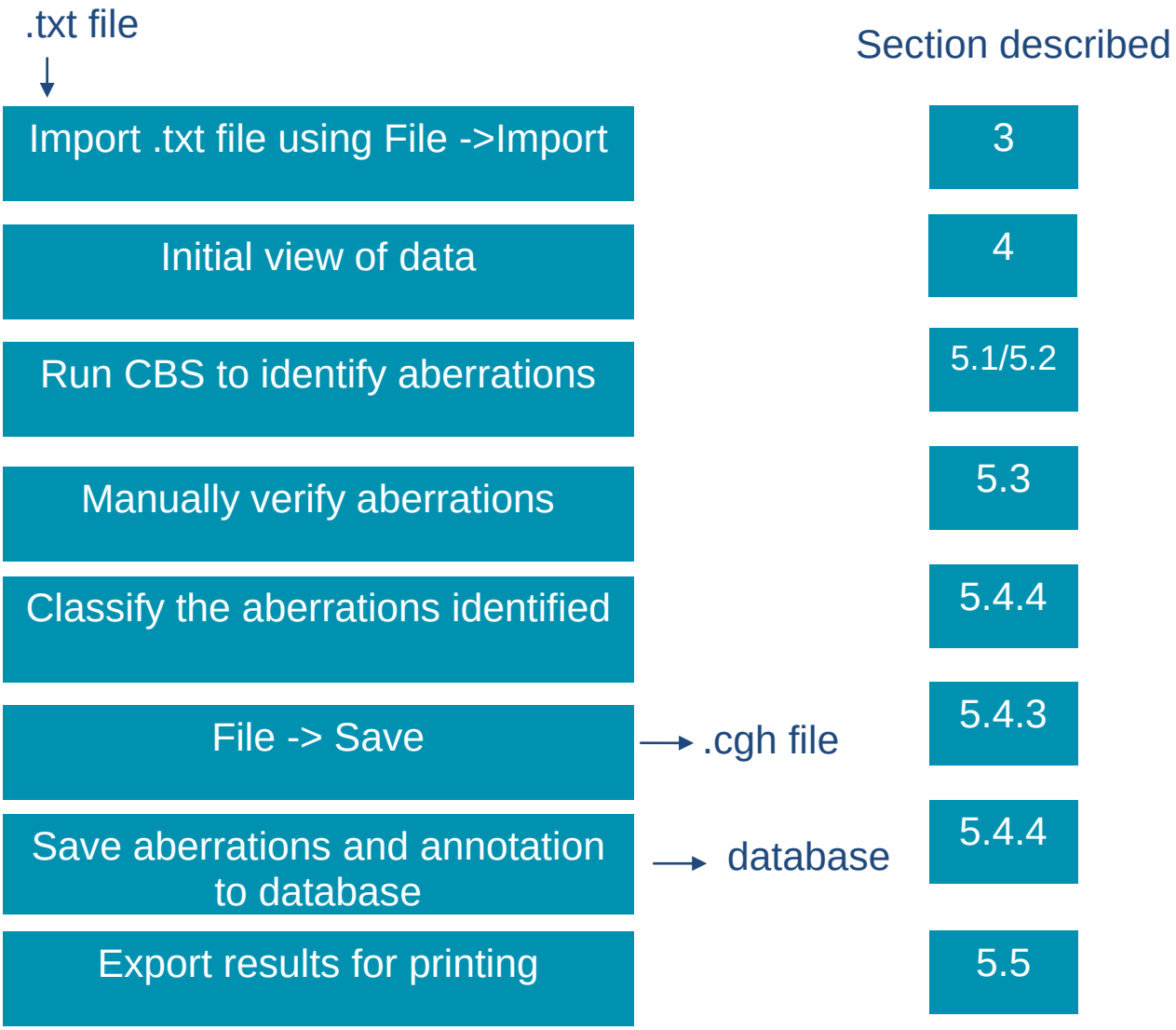
It is designed typically to support experiments where a reference experimental design has been carried out.

A reference experimental design is defined as follows:

Array 1: Sample 1 versus Reference
Array 2: Sample 2 versus Reference

CytoSure analysis software also supports a loop experimental design. For further details see section 6.

An overview of a typical analysis is shown in the flowchart below.



The software can be operated in two distinct modes.

Standard software mode: In this mode all the software options and settings can be altered as the user works through the analysis.

Workflow mode: Workflow mode simplifies data analysis and makes the routine analysis of aCGH data faster and more straightforward. There are two stages in the process. Firstly, protocol set-up and secondly, routine data analysis.

In the first instance, it is necessary to set up an analysis protocol. This task is best suited to more experienced users who can define and select the most appropriate settings for data analysis within their particular laboratory. The appropriate settings can then be saved. Routine users can then load the protocol and analyse the data using the previously defined settings. It is possible to alter protocol settings within Workflow mode however the software will not report that a protocol has been used in the reporting. The routine user will also be guided through the workflow outlined in the diagram above.
The setting up of the software in Workflow mode is covered in detail in Section 10 of the user manual. The use of the software in Workflow mode by a routine user is covered in Section 11.

1. Computer requirements, software installation and update

It is recommended that **Cyto**Sure Interpret software is installed on a computer with Windows XP or Vista. For proper display do not use Windows classic view.
For viewing and saving multiple datasets it is recommended that a computer with a Memory of 3 - 4GB is used.

To ensure fast running of the automated aberration detection it is recommended that the computer used has a quad core. If a quad core and multiple threads are used the processing time should typically be 4 to 5 minutes for Syndrome Plus 2x105k data.

CytoSure software works well on a non-quad computer; however the automated aberration detection running time will be slower.

To install the software double-click on the file *CytoSure Analysis software.exe* to run the installer and follow the on-screen instructions.

To update CytoSure analysis software uninstall the old version by selecting **Start** ->
All Programs -> **CytoSure Interpret software** -> **Uninstall.** Next install the most recent version as described above.

2. Running CytoSure Interpret software

Start the application via the Windows Start Menu. This will most likely be done by clicking **Start** ->
All Programs -> **CytoSure analysis software** -> **CytoSure analysis software**.

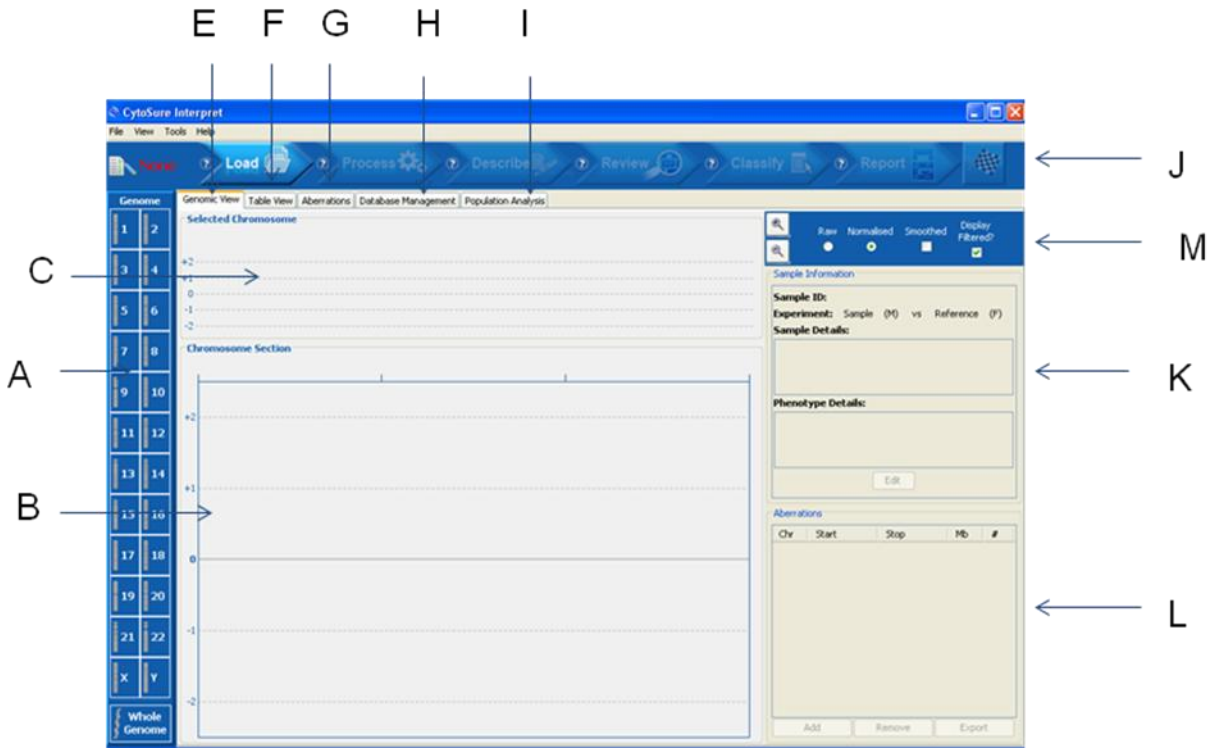


Figure 1 – Opening interface.

The opening interface is displayed in figure 1.

There are six parts to the interface:

A. Chromosome selection panel
This is used to select the chromosome whose data are displayed in the *chromosome overview* and the *chromosome section view*. It also offers the option to display data for the whole genome.

B. Chromosome overview
This panel displays a summary of the dataset selected in the *chromosome selection panel*, including a graphic of the chosen chromosome and a scatter plot of *log base 2* of the red/green mean signal ratio ($\log_2$ R/G) for each probe against its chromosomal/genomic base position. If the whole genome is selected, the panel indicates the chromosome boundaries in place of the chromosome graphic. Note that if the invert button is ticked then the data will be displayed as ($\log_2$ G/R). See section 4.1 for more details on inverting the data.

C. Chromosome section view
This panel plots a user-selected subset of the data available in the *chromosome overview* in the same manner ($\log_2$ R/G against chromosomal/genomic base position). When data are loaded, this panel also supplies the user with syndrome, gene and exon references with which plotted probes can be correlated as an aid to analysis.

D. Genomic view
The Genomic view displays the graph of the log ratio versus genomic position for the probes. There are also tracks showing annotation.

E. Table view
This panel displays detailed information about the probes that are visible in the chromosome section view. This includes the log ratio, red and green signal. See section 4.9.

G. Aberration tab
This panel displays details information about the aberration selected either by the user (manually) or using the automated aberration detection (see section 5). The aberrations can be annotated and edited (section 5.3)

H. Database management tab
The software has the ability to store the aberration results detected in a database. This tab allows the user to review and if necessary edit the database information.

I. Population Analysis
The software has the ability to analyse all the stored aberrations in the database. Users can analyse the frequency of aberrations across a population.

J. Workflow panel
The workflow feature is an easy use method of using the CytoSure software. It consist of a number of sequential steps, which are followed by clicking the buttons on the Workflow panel

K. Sample detail panel
Sample details and QC metrics and present here in this panel

L. Aberration panel
The aberrations called as a consequence of running the CBS or by calling manually are shown here.

M. Display adjustment toolbar
The toolbar provides a number of controls allowing the user to adjust the display to their requirements.

3. Loading data files

Opening feature extracted files

To display **Cyto**Sure data, the user must load a **feature extracted file**. The file should contain red and green signal values for each probe in the array being analysed. The software supports both **Agilent** and **Axon (GenePix)** format feature extracted files.

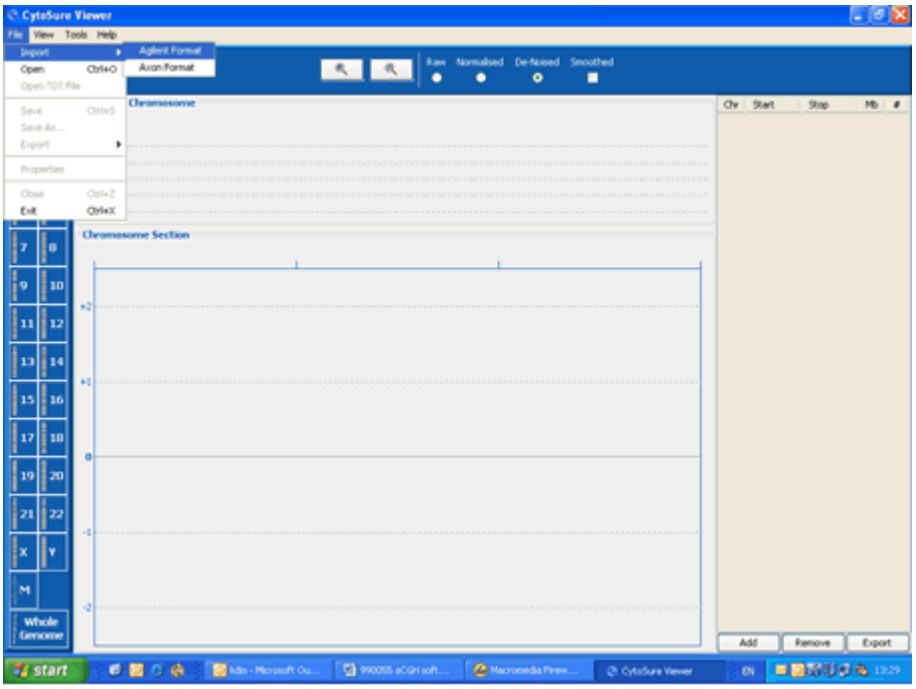


Figure 2 – Importing data files.

For analysis of mouse or rat files click File -> Select genome and select the relevant genome annotation. The default annotation is human.

To import a file, click **File** -> **Import**.

Select the feature extracted txt file you wish to import and click **Open**.

The sample details box appears (figure 3). Select which dye the sample was labelled with. The software can then calculate whether an aberration is a gain or a loss. In addition, select the sex of both the sample and the reference. Sample details and phenotype details can also be entered. It may also be necessary to select the array type (this will depend on array type used).

Click on the continue box.

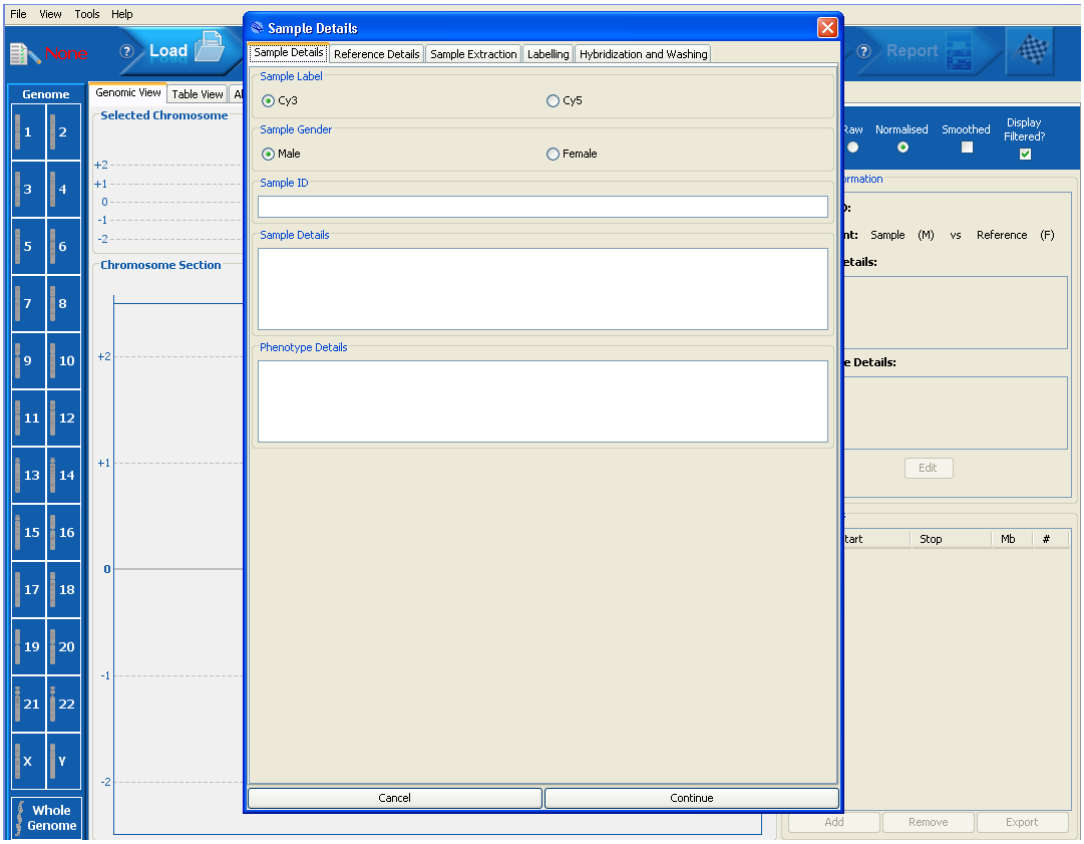


Figure 3 – Sample details box.

The software displays a normalisation box (see figure 4). CytoSure software incorporates a LOWESS normalisation algorithm. However, some feature extraction software (e.g. Agilent feature extraction) incorporates a normalisation option too. Therefore, if the data has already been normalised then click 'No'.

The recommended method is that the feature extracted data has not been normalised (see OGT technical note) and the user clicks 'Yes' to feature extract with CytoSure algorithm.

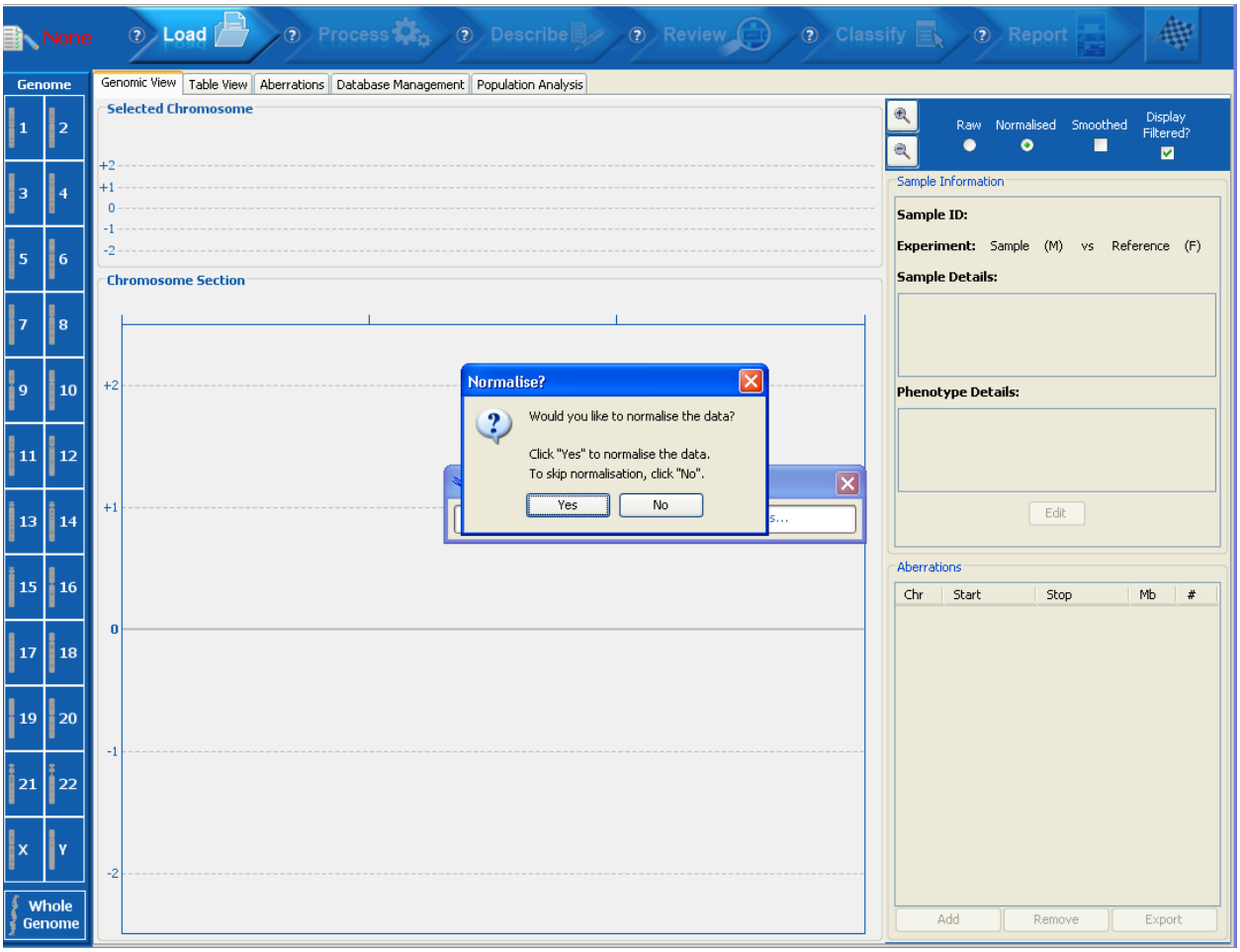


Figure 4 – The normalisation query box.

Loading in multiple files

Multiple files can be loaded in by sequentially using **File** -> **Import.** The different arrays results will be displayed in different colours. Note that data from more than 2 arrays may cause the computer to become slow, depending on the memory of the computer used.

When there a multiple datasets loaded, a series of tabs will appear on the right hand side of the screen. Clicking on the tabs will toggle between the sample details of each dataset. Associated with each tab are some tick boxes. These are shown in the figure below.

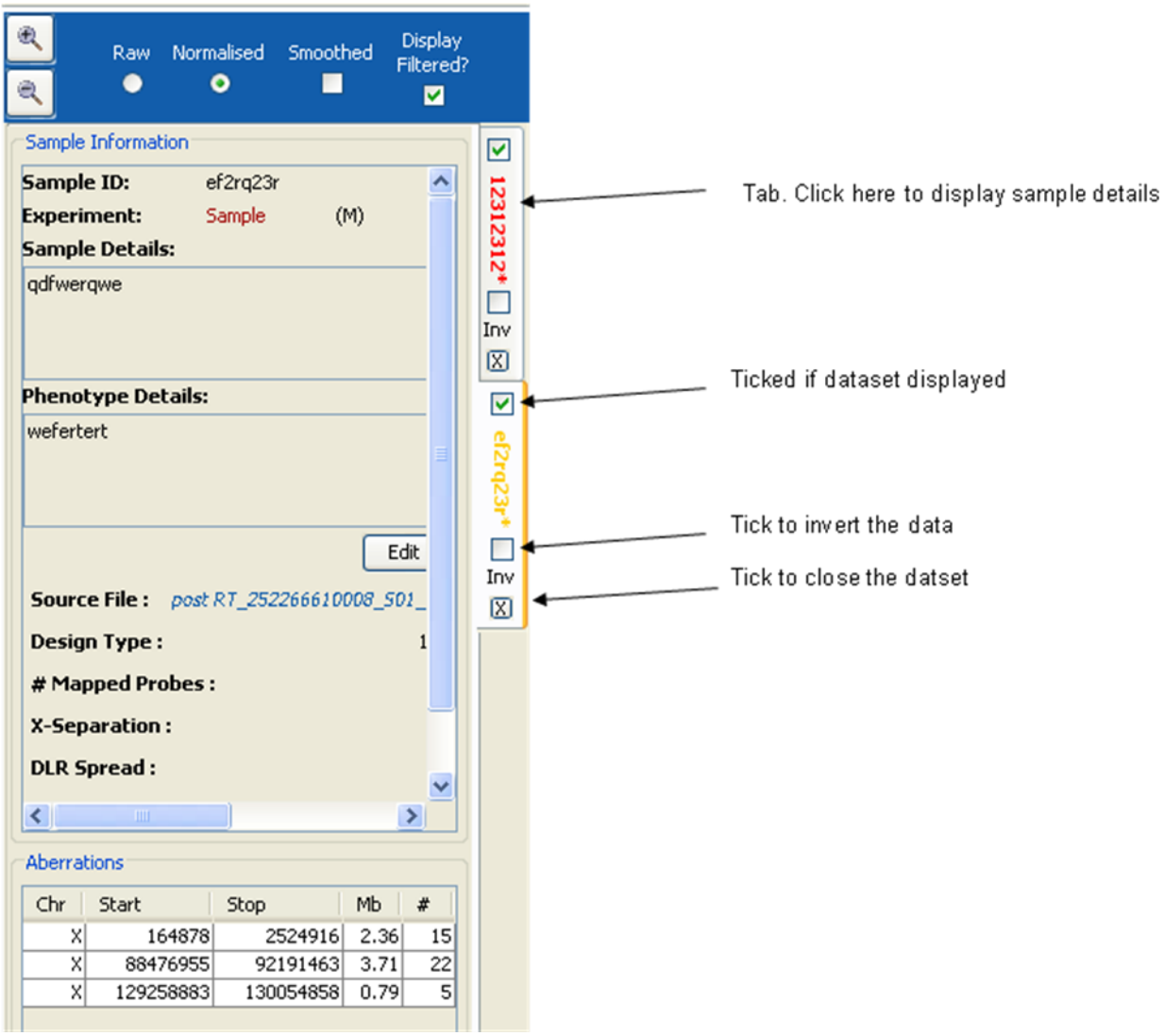


Figure 5 – The tabs and the radio buttons

Opening previously analysed files
Files that have been previously analysed using **CytoSure analysis software** and saved can be reopened in **CytoSure analysis software** by using **File** -> **Open**

4. Viewing CytoSure data

Once data files have been loaded, the user is able to browse the data using a number of controls supplied by each component of the interface.

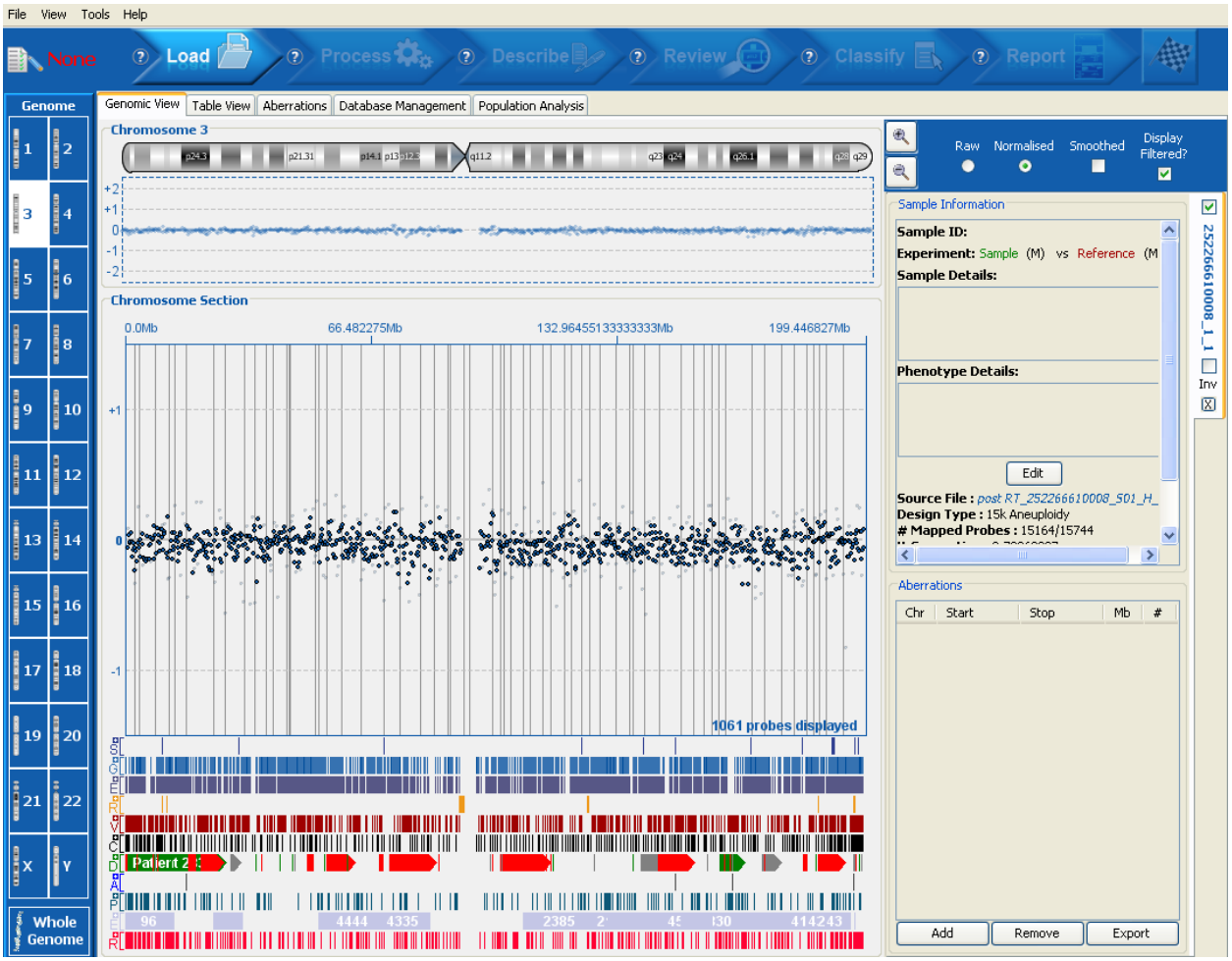


Figure 6 – The viewer with feature extracted files loaded.

4.1. Navigation

The **chromosome selection panel (A)** displays an icon for each chromosome in the Human Genome. This allows the user to view the data for a specific chromosome. At the bottom of the

panel is a button entitled “Whole Genome”; this allows the user to view all the data available.

To view a specific chromosome dataset, click its corresponding button.

- A chromosome selection button. Click on this button to display data for chromosome 10. The M button refers to Mitochondrial DNA

- Click on this button to display data for the whole genome

- A disabled chromosome selection button. There are no probes mapped to this chromosome in the current dataset

The **chromosome overview** (B) displays a summary of the dataset selected in the **chromosome selection panel** (A). If the data has been **smoothed**, this component also shows a line indicating the smoothed values of the displayed data. Initially, the scatter plot is outlined with a dashed rectangle indicating the current section of the chromosome (or genome) that is displayed in the **chromosome section view** (C).

The user can select a subsection of the chromosomal/genomic data by dragging the mouse between the required start and stop base positions of the chromosomal/genomic data. The **chromosome section view** (C) will then be updated accordingly. If the mouse is positioned within the **chromosome overview** (B), the user is also able to scroll through the chromosome/genome whilst maintaining the size of the selection window by using the mouse wheel or the arrow keys.

Chromosome section view (C) displays the data subset selected in the **chromosome overview** (B) as a scatter plot of probe location against $\log_2$ red/green mean signal ratio. The user is able to view information about each displayed probe by hovering over the probe with the mouse.

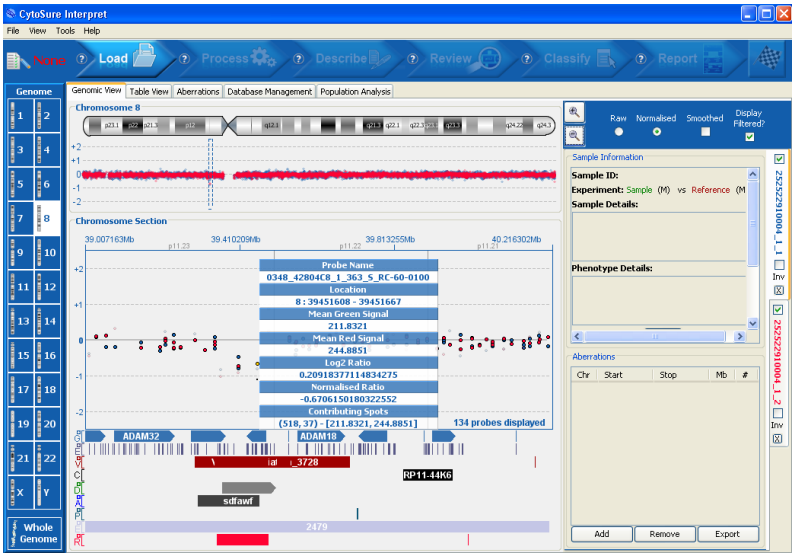


Figure 7 – Hovering over a probe

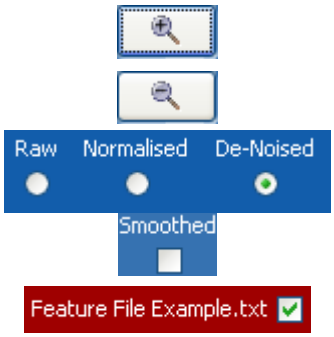
The following information is displayed:

- name of the probe
- location of the probe within its associated chromosome
- mean red and green signal values
- $\log_2$ ratio of these values
- any other calculated ratios, if available
- individual signal values for each spot on the array contributing to the mean signal values

The user is able to navigate the chromosome/genome in the same manner as in the **chromosome overview** (B), with the addition of being able to traverse the signal ratio axis as well as the location axis. This is actioned by using the corresponding arrow keys, or the mouse wheel with the *Alt* key pressed. The user can also alter the size of the signal ratio axis by pressing the *Shift* key whilst scrolling with the mouse wheel, or pressing the up or down arrow keys.

Hovering over a reference will display the information for that reference, and clicking on the reference will open a browser window and navigate to a relevant web page, if the user's internet connection allows.

The **display adjustment toolbar** (M) provides the following controls:



Zoom in

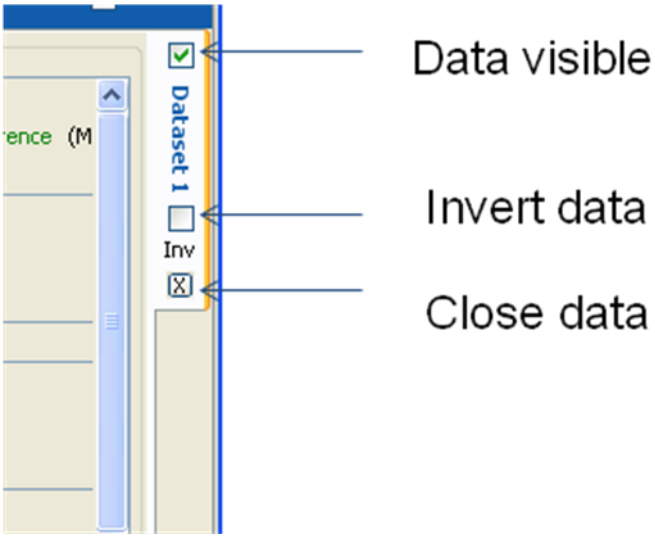
Zoom out

Toggle between displaying the raw, calculated normalised data or denoised data

Select the calculated smoothed data

Select a particular dataset (when at least one has been loaded)

The invert tick box will flip the relevant dataset. The standard default (unchecked) displays the data with the y-axis as log2 (red/green). With the box checked the particular dataset will be displayed with a y-axis value of log2 (green/red



The Tab at the right hand of the screen contains various buttons.

The invert tick box will flip the relevant dataset. The standard default (unchecked) displays the data with the y-axis as log2 (red/green). With the box checked the particular dataset will be displayed with a y-axis value of log2 (green/red)

The Data visible tick box, will turn on and off the display of the data

The close data button will remove the data from the CytoSure software following a warning

At the bottom of the **chromosome section view** (C) is the annotation. Further details are supplied in section 4.2. Finally the cytogenetic location is displayed with the grey lines.

Whole Genome view

Clicking on the

will display the whole genome view. Clicking on the

button will then display the whole genome, where the probes are plotted in sequential order rather than their genomic position. In this view it is often easier to locate aberrations.

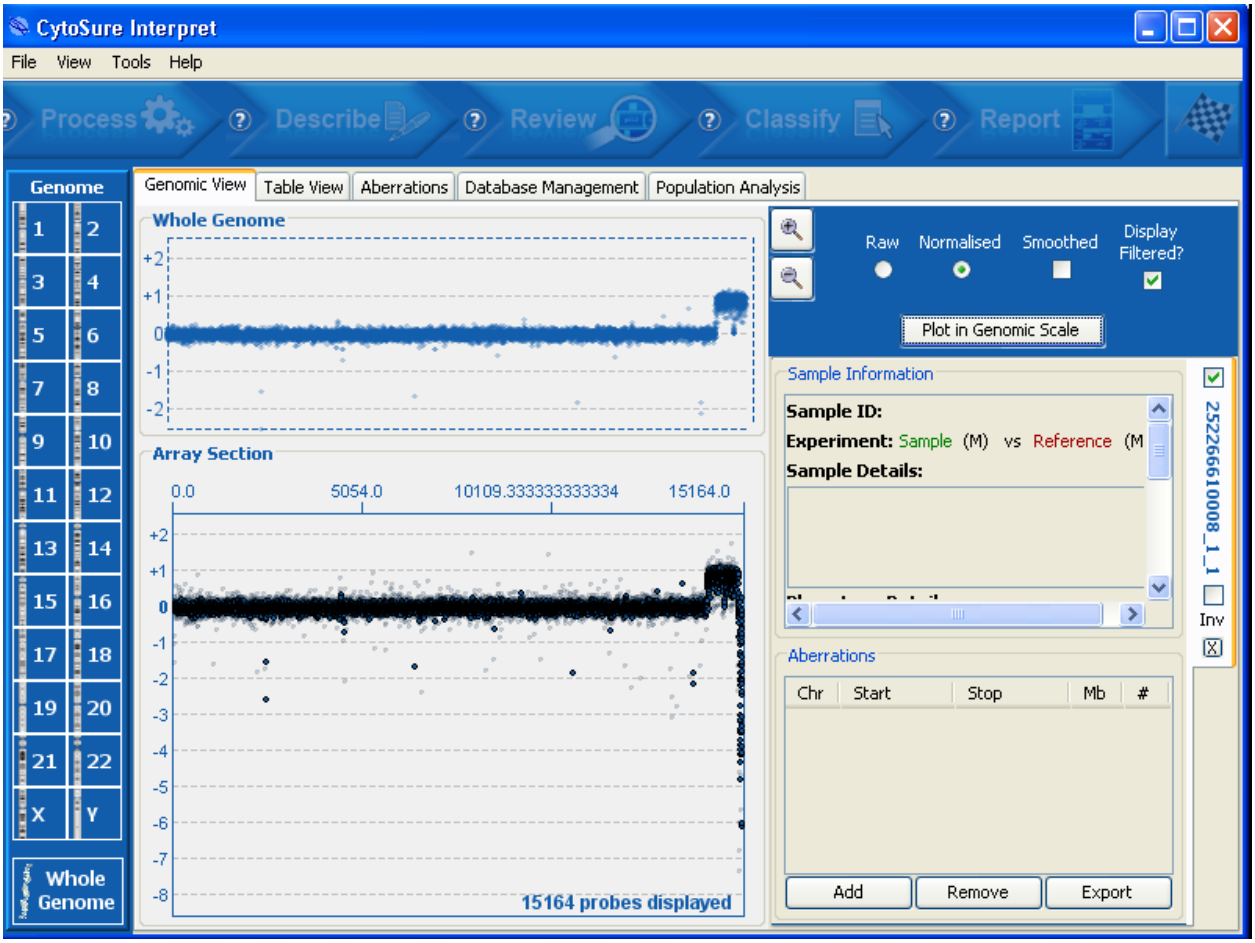


Figure 9 – Whole genome, equally spaced view.

4.2 Annotation

The **chromosome section view** (C) also provides the user with a number of references to regions of interest within the selected section. References are grouped into:

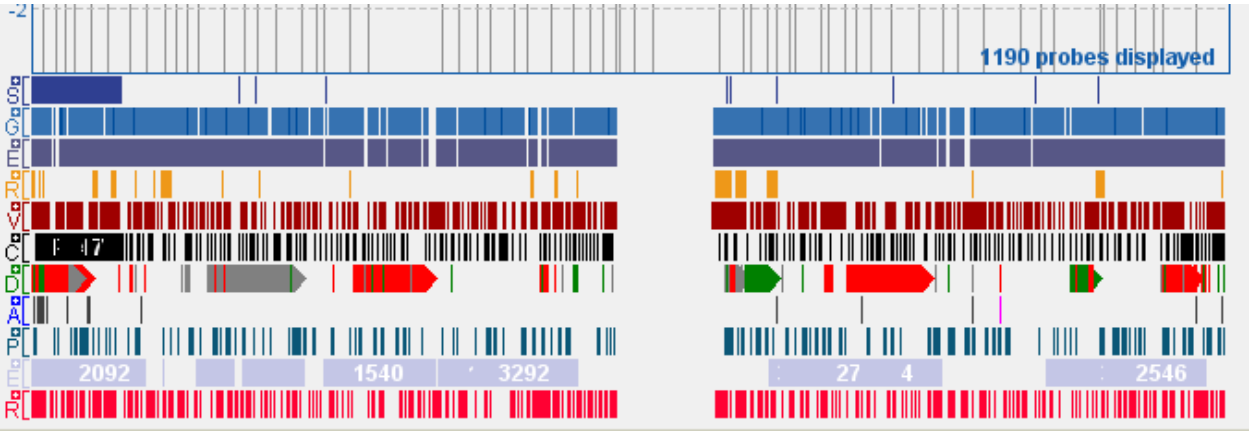


Figure 10 – Annotation tracks.

Syndromes (S) dark blue: Some common Syndromic areas have been included. Note this is not exhaustive and it is only an indication that this is a Syndromic region. Clicking on these bars will take the user to the relevant OMIM entry.

Genes (G) light blue: Source - Ensembl annotation. Left click on the bars to link to the relevant Ensembl entry.

Right click on the bar to display the following options:

- Link to Ensembl
- Link to UCSC
- Link to iHOP (information hyperlinked over proteins) - a network of concurring genes and proteins extending through the scientific literature touching on phenotypes, pathologies and gene function
- Link to Genecards - Information about the gene
- Link to GeneRIF – Information about the gene
- Link to Prospectr – Prospectr **theoretically** calculates if a gene is a disease gene
- Link to WikiGene - Information about the gene

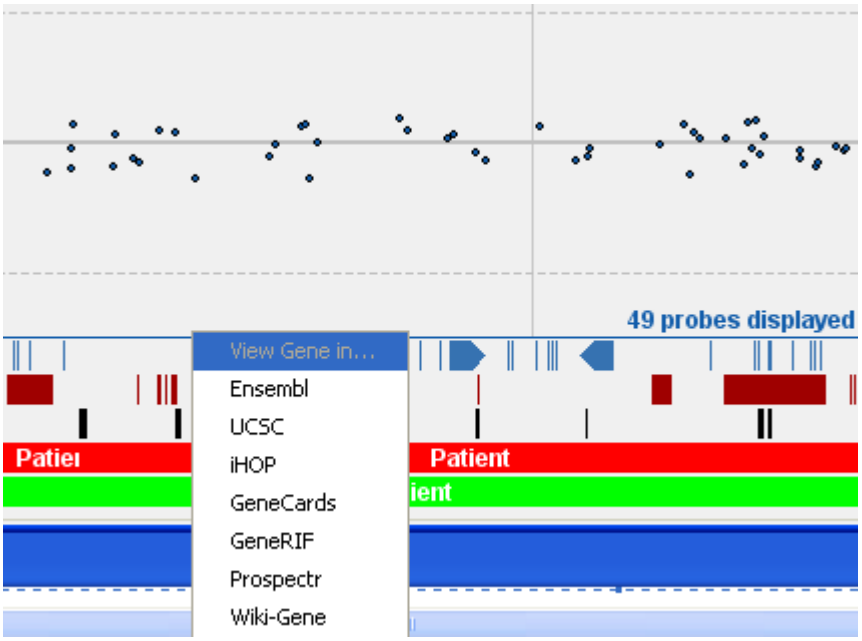


Figure 11 – Links for further information on genes.

Exons (E) purple: Some genes have the position of the exons included. Clicking on these bars will take the user to the relevant Ensembl entry.

Recombination hotspots surrounded by segmental duplications (D) yellow: Regions of the genome defined as a recombination hotspot by Bailey *et al* Science **297** 1003-1007. These regions are typically surrounded by segmental duplications.

Copy number variations (V) dark red: Source - Toronto DGV database. This extensive database shows the position of human CNVs. There may be errors in the Toronto database, so these regions should only be regarded as an indication that there is a benign CNV. Clicking on these bars will take the user to the relevant DGV entry.

Confirmation bar (C) black: This displays the position of various publically available FISH / BAC and MLPA probes that can be used for confirmation.

Database track (A) blue: This shows the positions of aberrations recorded by previous experiments which have been saved in the database

Double clicking on the bar will display an image and annotation of the aberration that has been saved to the database (figure 10).

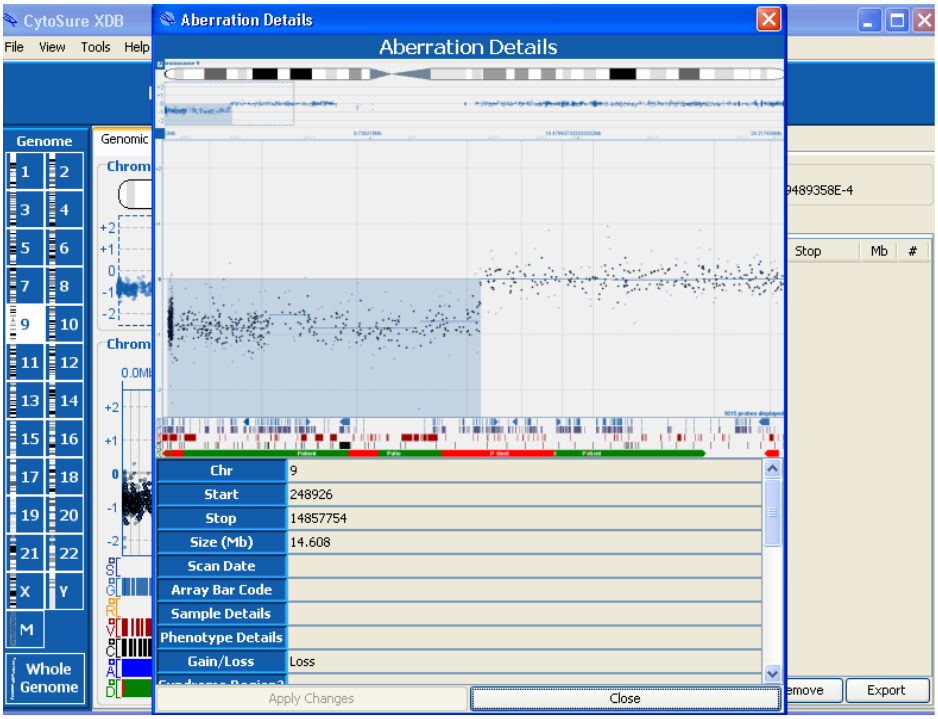


Figure 12 - Aberration details of an aberration that has been saved to the database.

Decipher (D) red / green: Position of patient deletions (Red) and duplications (Green) from the Decipher database (https://**decipher**.sanger.ac.uk/). Click on the bar to access the relevant Decipher page. For up to date information please access the Decipher web site directly.

CNV data (P). *CNV* data from Shaikh *et al* Genome Research http://genome.cshlp.org/content/early/2009/07/10/gr.083501.108.abstract

By default this track is turned off. To turn on the track see the section on 'adding back annotation tracks' below. Having many tracks open can slow the software.

This extensive study analyzed 2,026 disease-free people using an Illumina microarray platform. The population was mainly Caucasians or African Americans. The data downloaded is the CNV block data (please see http://cnv.chop.edu/help.jsp;jsessionid=EE8A0897783507053E52A8A2E185C076?sec=cnv_view#cnv_view for the definition of a CNV block).

Included when the user hovers over the annotation is the % frequency that a particular CNV block is present within the population studied and the number of CNVs within that block which are gains or losses. Data is supplied courtesy Center for Biomedical Informatics at the Children's Hospital of Philadelphia (http://stokes.chop.edu/cbmi)

Redon CNV data (R). CNV data from http://www.sanger.ac.uk/humgen/cnv/42mio/

By default this track is turned off. To turn on the track see the section on 'adding back annotation tracks' below. Having many tracks open can slow the software.

This study looked at 41 samples at very high resolution using multiple arrays with a total of 42million probes. Unfortunately there is no gain or loss annotation.

ECARUCA track (E). Light Blue. Data from the ECARUCA database of rare chromosomal aberrations. For more details on the particular aberrations, please visit http://agserver01.azn.nl:8080/ecaruca/ecaruca.jsp. Unfortunately there is no direct link to Ecaruca entries.

By default this track is turned off. To turn on the track see the section on 'adding back annotation tracks' below. Having many tracks open can slow the software.

Some of the annotation tracks can be filtered to only display certain information. Please see section below on annotation filtering.

Expanding the annotation bars

To view the full annotation, click on the button shown in the diagram below:

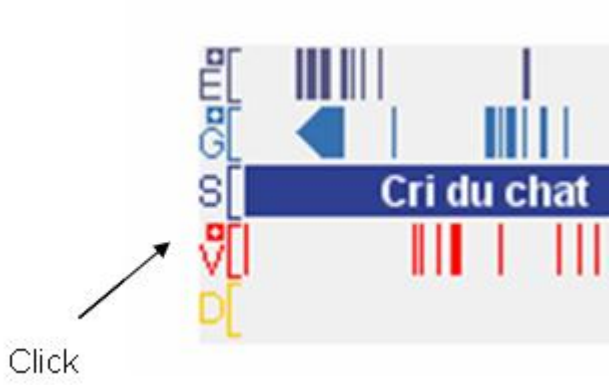


Figure 11 – Method to enlarge the annotation bars

Full CNV annotation expanded

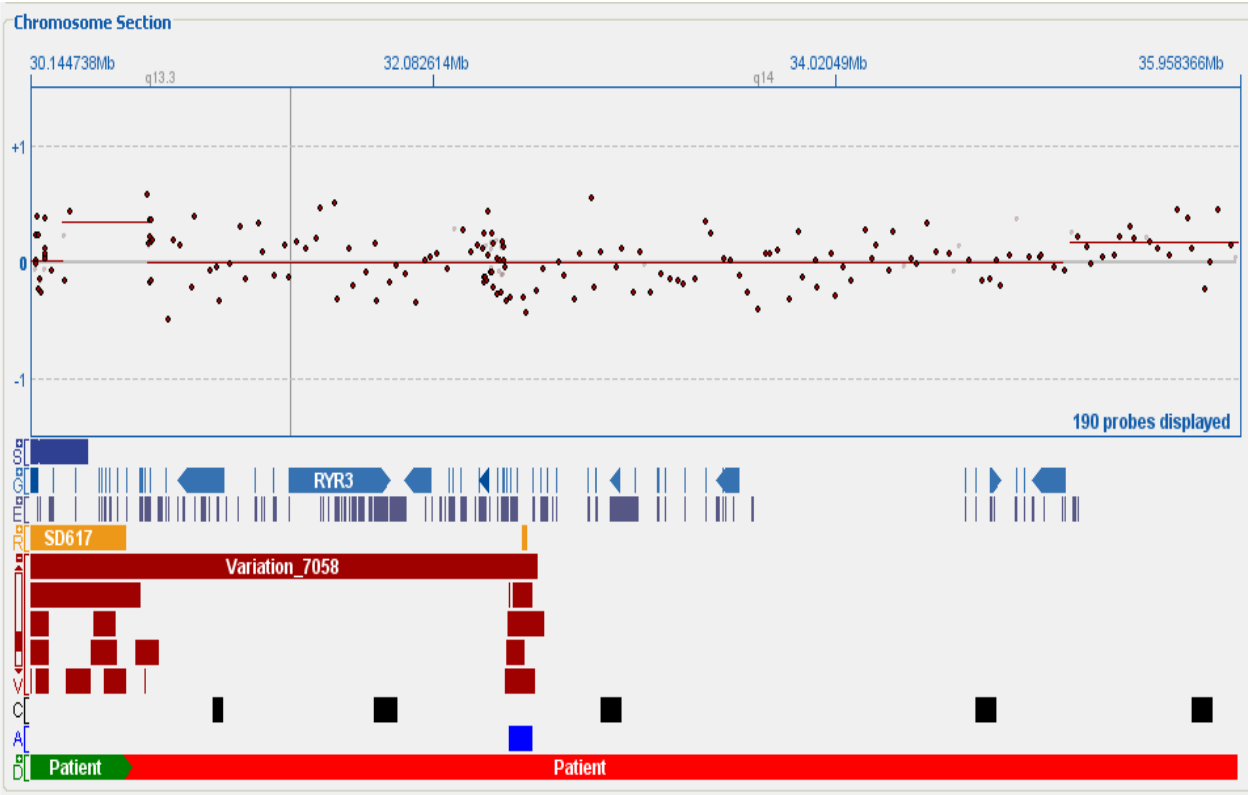


Figure 13 – CNV annotation displayed.

The annotation is displayed (figure 13), and in order to view all the annotation the slider bar can be used to scroll through the annotation (figure 14)

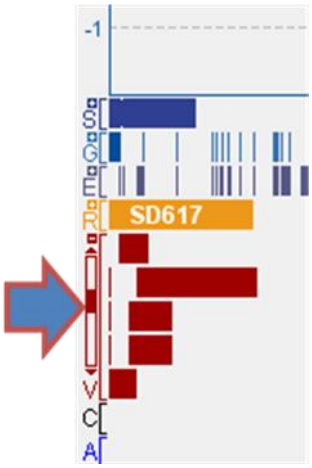


Figure 14 – View of the slider-enables scrolling through the annotation

To summarise the data, a frequency plot can be plotted. Right click in the relevant annotation track and select the option ‘show frequency plot figure 15). The frequency plot is a graph where the y-axis is the frequency of entries in the database (for CNVs the database is the DGV). The plot is shown in figure 16.

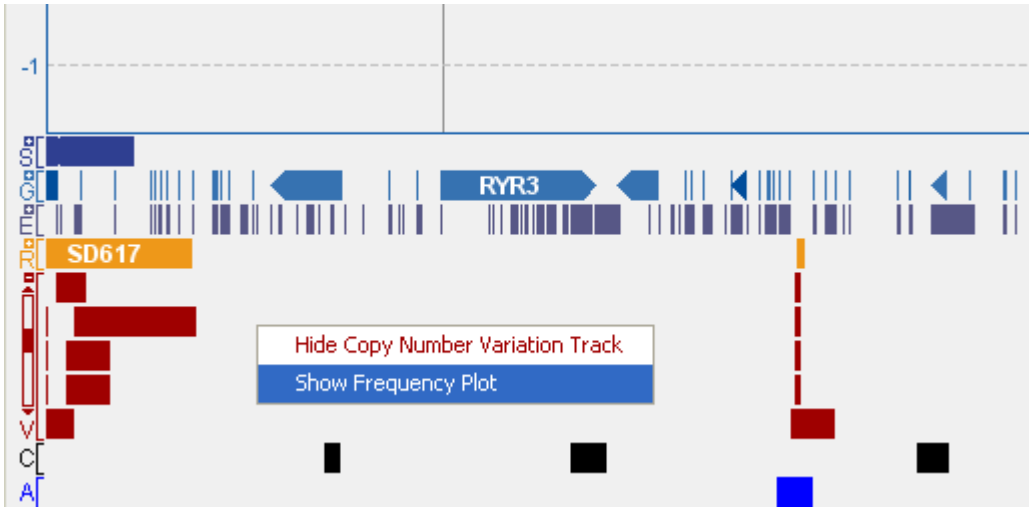


Figure 15 - To select the frequency plot, right click in the annotation track and select the show frequency plot option.

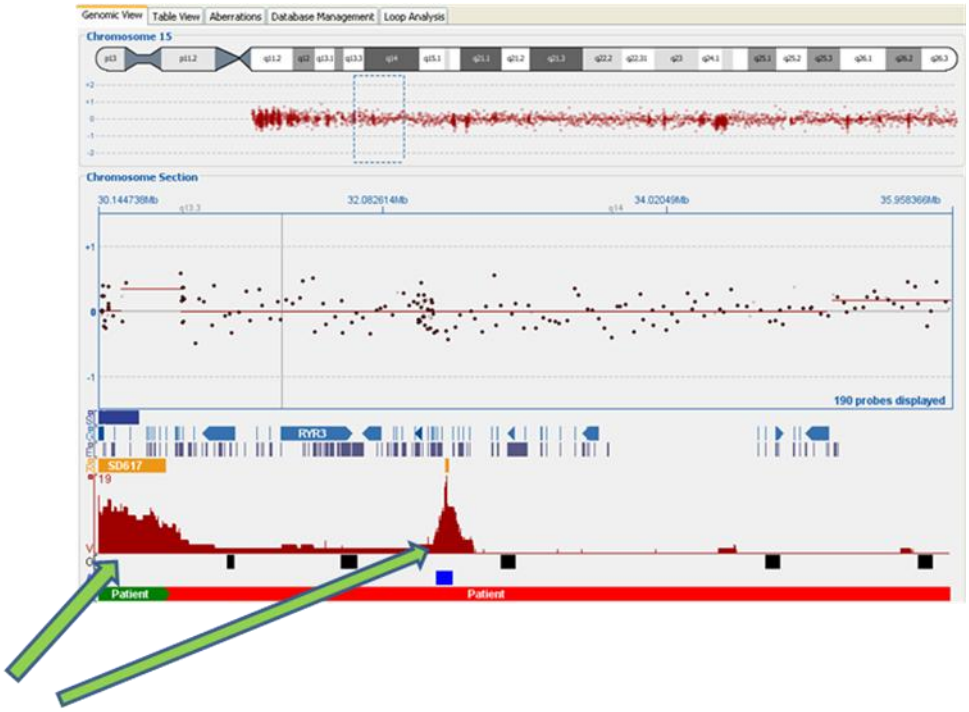
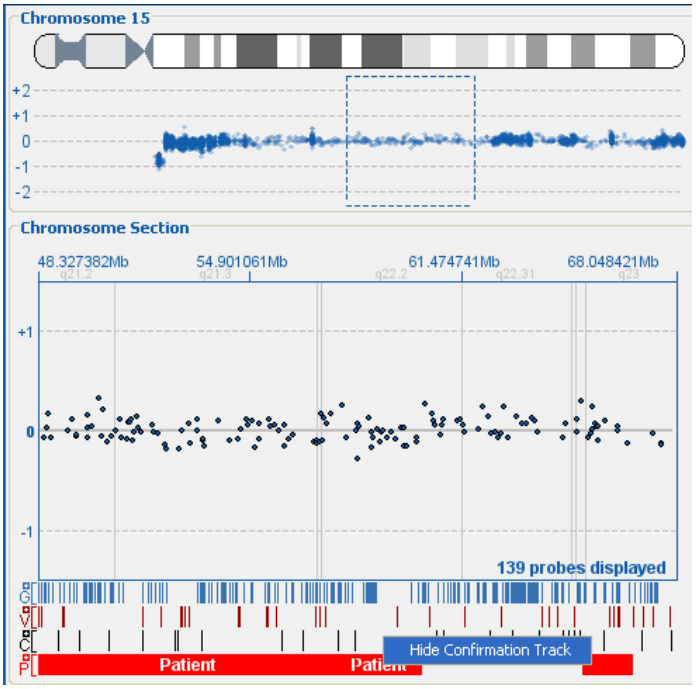


Figure 16 – View of the frequency plot.



Removing the annotation tracks

To remove an annotation track, right click in the relevant track outside a bar (see diagram below). A pop-up box will appear, “Hide confirmation track”. Click this to remove the relevant track.

Figure 17 – Hiding tracks.

Adding back annotation tracks

Users can set up their own annotation tracks containing for example data obtained on BAC arrays or by Karyotyping. To add custom annotation tracks select Tools -> Options -> Annotation

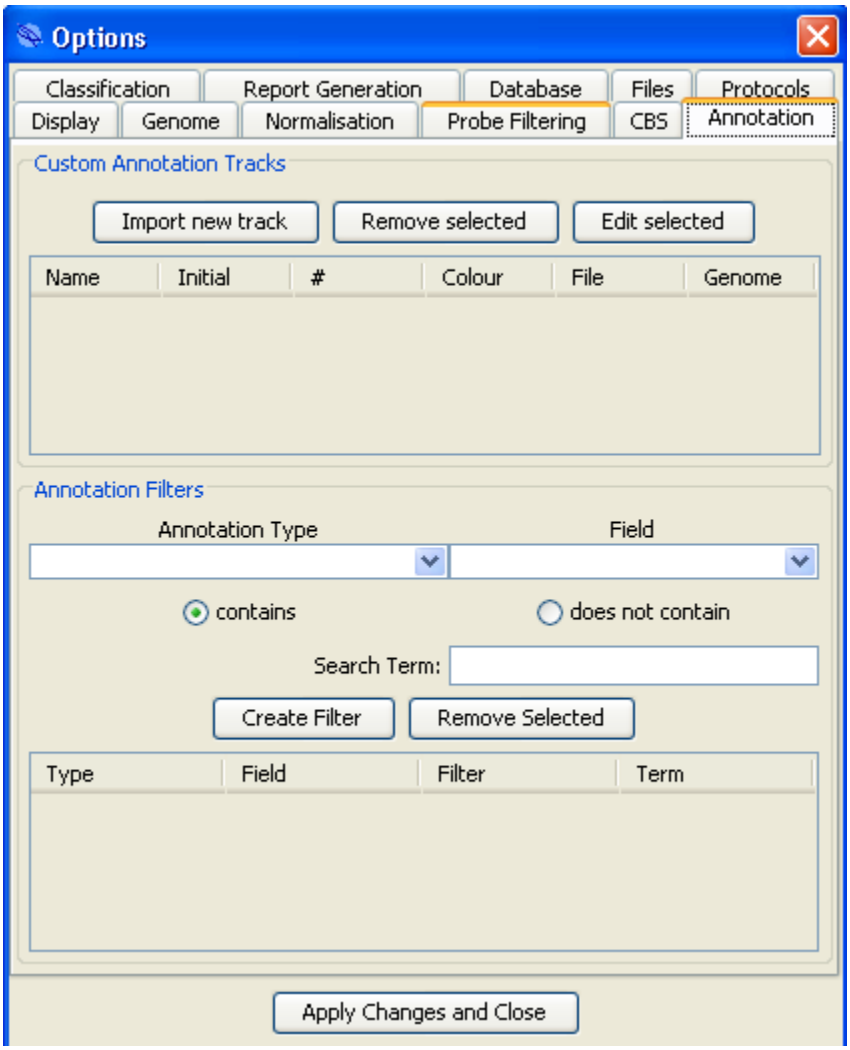


Figure 18 – Adding custom annotation

To add custom annotation, prepare a .txt using Microsoft Excel. The file needs to have columns containing the following data: Chromosome, Start, Stop and Number (or Sample ID). An image of an appropriate .txt file opened in Excel is shown below

A	B	C	D	E
Number	Chromosome	Start Position(bp)	End Position(bp)	Arraytype
106	1	918164	1145847	1Mb Clone Array
129	1	175311915	187408224	Tilepath cloneset
218	1	3214534	6042798	1Mb Clone Array
263	1	145434722	146342686	1Mb Clone Array
611	1	219787706	221890504	1Mb Clone Array
750	1	190142199	190333427	1Mb Clone Array
881	1	105468369	108512967	1Mb Clone Array
890	1	145434722	146342686	1Mb Clone Array
894	1	941583	7843532	1Mb Clone Array
983	1	144149973	146342686	1Mb Clone Array
991	1	216532635	218866017	1Mb Clone Array
1015	1	218983622	224229492	Tilepath cloneset
1186	1	145434722	146342686	1Mb Clone Array
1389	1	144979554	146204124	Tilepath cloneset
1458	1	246013826	246181016	1Mb Clone Array
1575	1	108330015	111090422	1Mb Clone Array
1588	1	215209972	225807728	1Mb Clone Array
1617	1	120270573	121186957	WTSI-TPA
1634	1	20428363	23311475	1Mb Clone Array
1803	1	18255166	19752047	Cytochip_V1.1
1925	1	144807807	146375718	Agilent - Human 244
1927	1	69875752	70814713	1Mb Clone Array

Figure 19 – Adding an annotation track

Click ‘Import New track’ and navigate to the .txt file. The box below opens

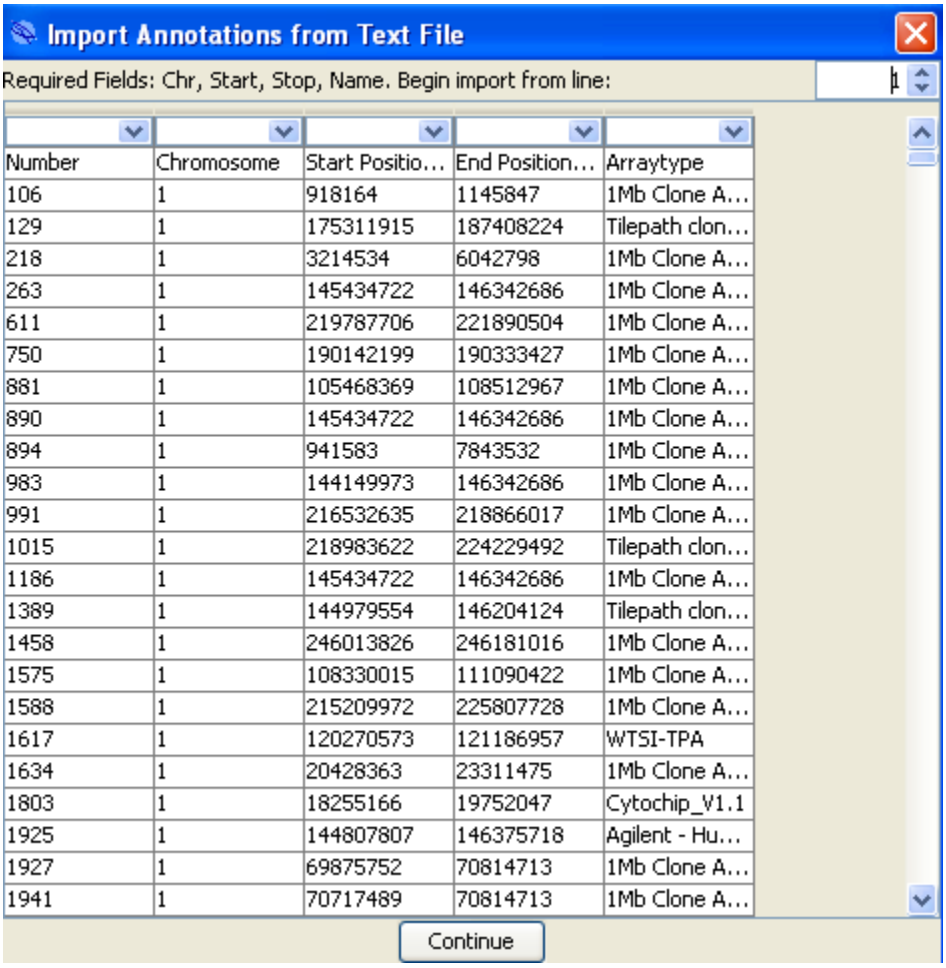


Figure 20 – Adding an annotation track

The user needs to input where the data begins. In this case the data starts at Line 2 as the first line contains the headers. Next the user needs to define the type of data present in each column. To do this, select the appropriate option from the drop down menu at the top of each column. In this case the first column is Name. The second column is Chr, the third column is Start and so on.

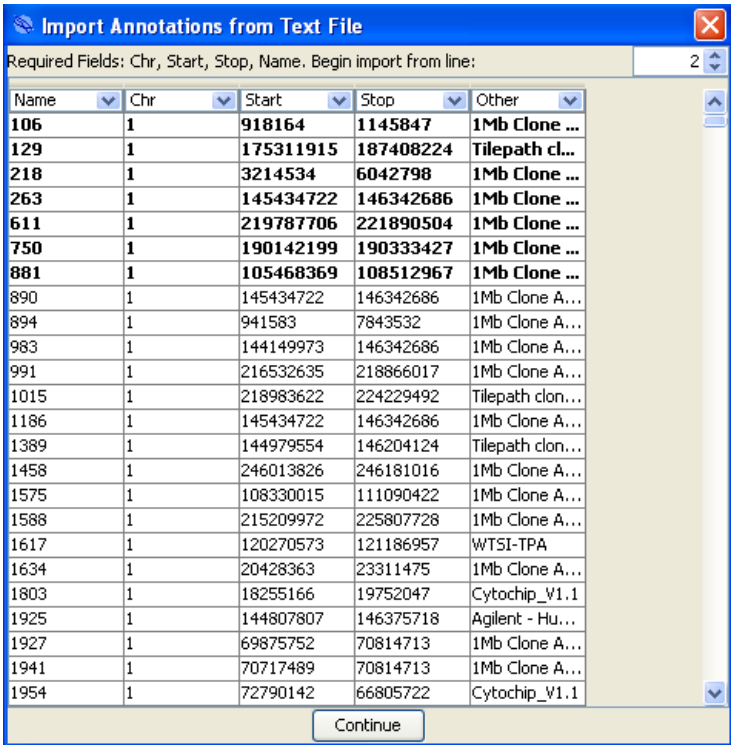


Figure 21 – Adding an annotation track

Note that URLs can also be included. Click ‘continue’. Within Custom Annotation Track Details, it is possible to assign a name, an identifying initial and a colour to each track. Once selected, the annotation will be visible in the Custom Annotation Table. In the genomic view the annotation will also be visible in its own annotation track.

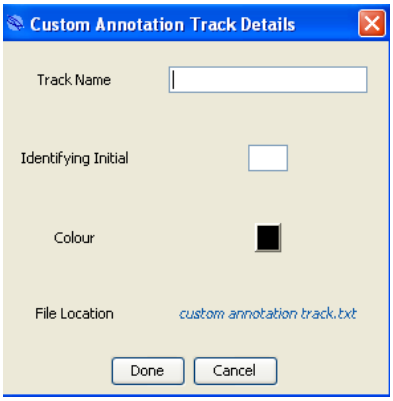


Figure 22 – Track details

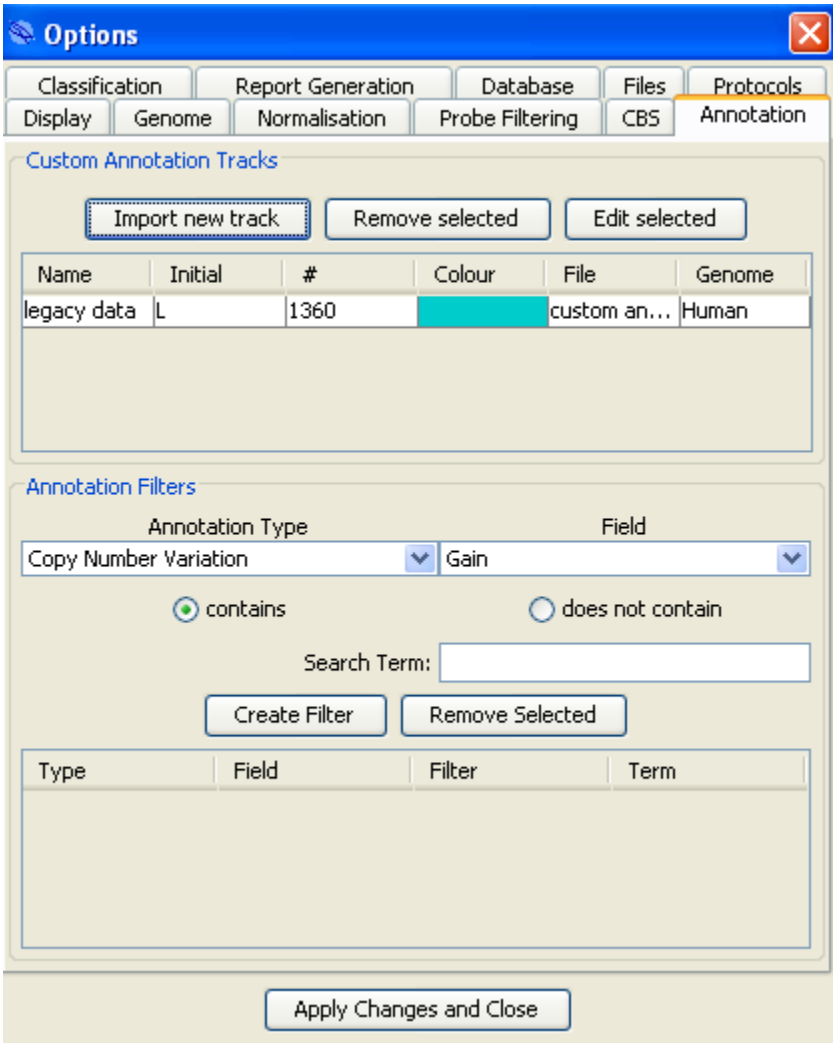


Figure 23 – Adding an annotation track

Filtering the Annotation

The ability to filter an annotation track has been included in this software release. This functionality now enables users to select annotation data for removal. For example in the DGV/CNV track the user might wish to remove CNVs that are inversions. Alternatively the user may wish to avoid having BAC data in the track,

To filter an annotation track select Tools -> Options -> Annotation

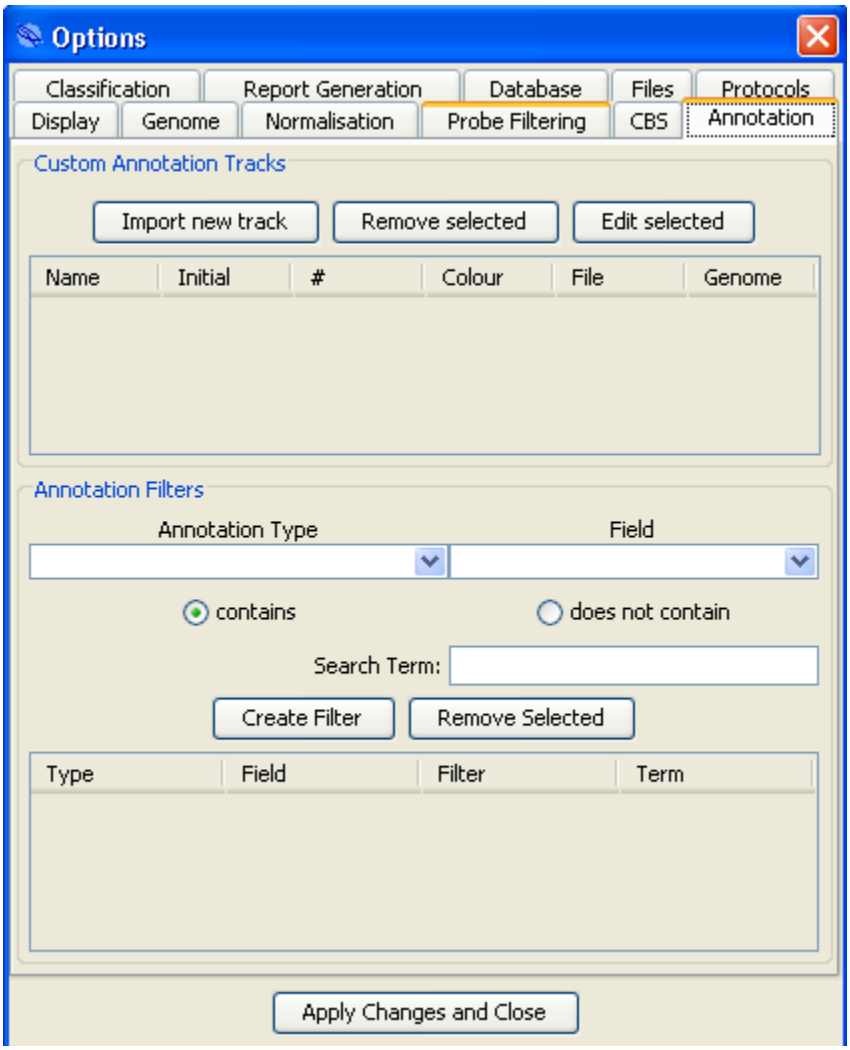


Figure 25 – Select annotation track

Select the Annotation Type to be filtered. In the example shown in the figure below, the 'Copy Number Variation' Track has been selected from the 'Annotation Type' drop-down menu. Then using the drop down menu click the field (or column) where the data is going to be filtered. In this example, in order to remove CNVs that are Inversions, select the field 'Type' and then select the radio button 'does not contain' and type 'inversion' in the search term box

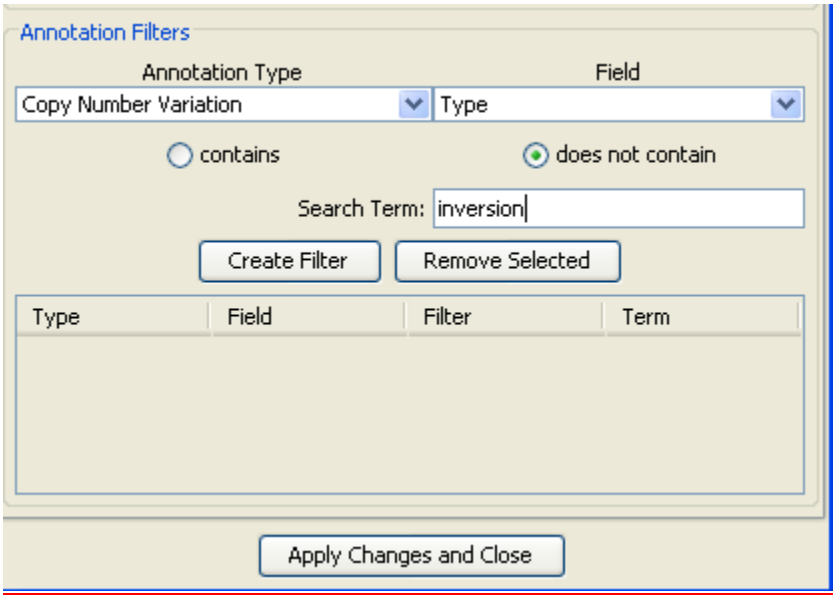


Figure 26: This will remove all CNVs from the annotation that have been classified as inversions

Finally select 'Create Filter' and 'Apply Changes and Close'

4.3 Searching for a gene or probe or syndrome

- A probe - by clicking **View** -> **Find Probe…**, typing the name of the probe in the dialog that appears and clicking 'Ok'
- A gene or syndrome – by clicking **View** -> **Find Annotation …**, typing the name of the reference in the dialog that appears and clicking 'Ok'. The text needs to match the Ensembl annotation or the syndrome annotation used in the software. If not found try a synonym.
- A position in the genome – by clicking **View -> Go to position (bp)…,** , typing the number in the dialog box and clicking 'Ok'

4.4 Data display (the radio buttons)

There are 4 radio buttons on the Display adjustment toolbar (D). This allows the user to toggle between different views of the data

- Raw button – shows the data un-normalised
- Normalised button – shows the data after normalization.
- Smoothed button – see section 4.6
- Filtered – If checked the filtered probes are displayed. If unchecked all data points are shown, however the filtered probes are shown with lower brightness. The filtration options are discussed below in section 4.5.

4.5 Filtration of data

Filtration of the data is when certain data points are removed on the basis of certain parameters which are set by the user. For example data points which have low signal intensities on the array might be excluded from the analysis.

The filtration options can be accessed by clicking Tools -> Filters. Alternatively they can be accessed using Tools >Options > Filters. A box appears which gives the user various options (see figure 26).

Figure 26 - Filtration options.

The options available are as follows:

Absolute normalised log ratio – allows the user to filter the data so that only data points within a range of log ratio values are included

Signal to noise ratio – the SNR is calculated and the range of data points that are included can be typed into the box for both green and red signals

Signal intensity – the signal intensity range that can be included are typed into the box for both the green and red signals. These values are going to vary according to which scanner is used. However, as a guide line for Agilent scanners probes with signals of below 150 and above 50,000 should be excluded.

Exclude Non-Uniform Outliers – these are spots which have been flagged as outliers by the Agilent feature extraction software. If this box is ticked then these spots are removed from the analysis

Denoise – This is a form of outlier removal in which a percentage of outliers from a set number of probes are removed. The default value is 10% outliers from a window of 20 probes. So, in this case the 2 probes with the highest and 2 probes at the lowest ratios are not shown out of a window of 20 probes. The full dataset (non de-noised) is used for any calculations made by CytoSure analysis software.

To adjust default value click **Tools** -> **Options -> Filters**, and alter % outliers in pop-up box.

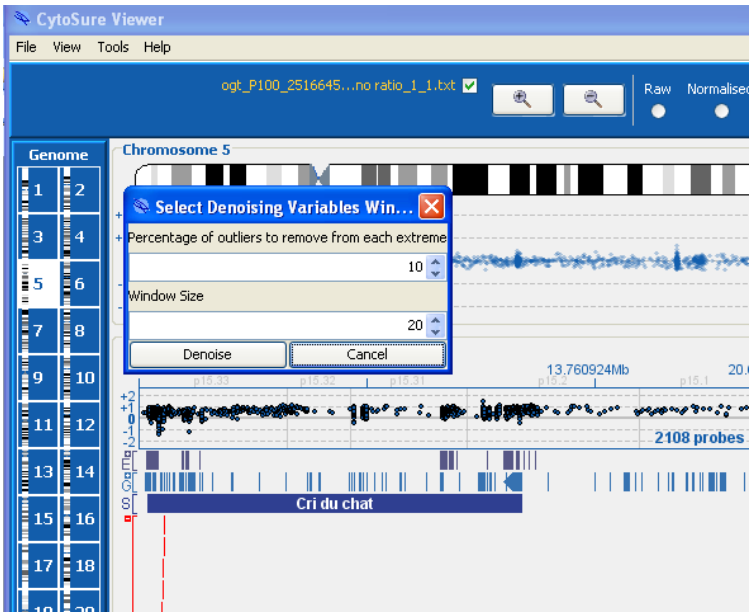


Figure 27 – Adjustment of outlier removal values.

Currently the non-filtered data is used for the CBS analysis. The ability to analyse filtered data will be available in future CytoSure software versions.

4.6 Normalisation and smoothing

Normalisation

On loading the data into CytoSure, the data is automatically normalised by the LOWESS algorithm incorporated into CytoSure. It is possible to alter the normalisation window size by **Tools** -> **Options** and then clicking on the **Tools** tab**.**

Specific normalisation

Particularly with custom arrays where a majority of probes may consist of an aberration, the LOWESS normalisation may not work correctly. There is an option to use specific normalisation. In this, use the navigation tools to display a probe subset in panel C that will act as control probes for normalisation. Select **Tools** -> **Normalise on the current subset**. The software will recalculate the normalisation using the probes in the display.

This feature is unlikely to be required for the analysis of data from Syndrome Plus arrays.

Smoothing

To smooth the array data, select **Tools** -> **Smooth Data**, choose a smoothing window size in the dialog box that appears, and click “Smooth”. A larger window size will result in a smoother line through the data. The window size number is in base pairs. Once the software has calculated the smoothing line, click on the Smoothed radio button to display the Smoothed line.

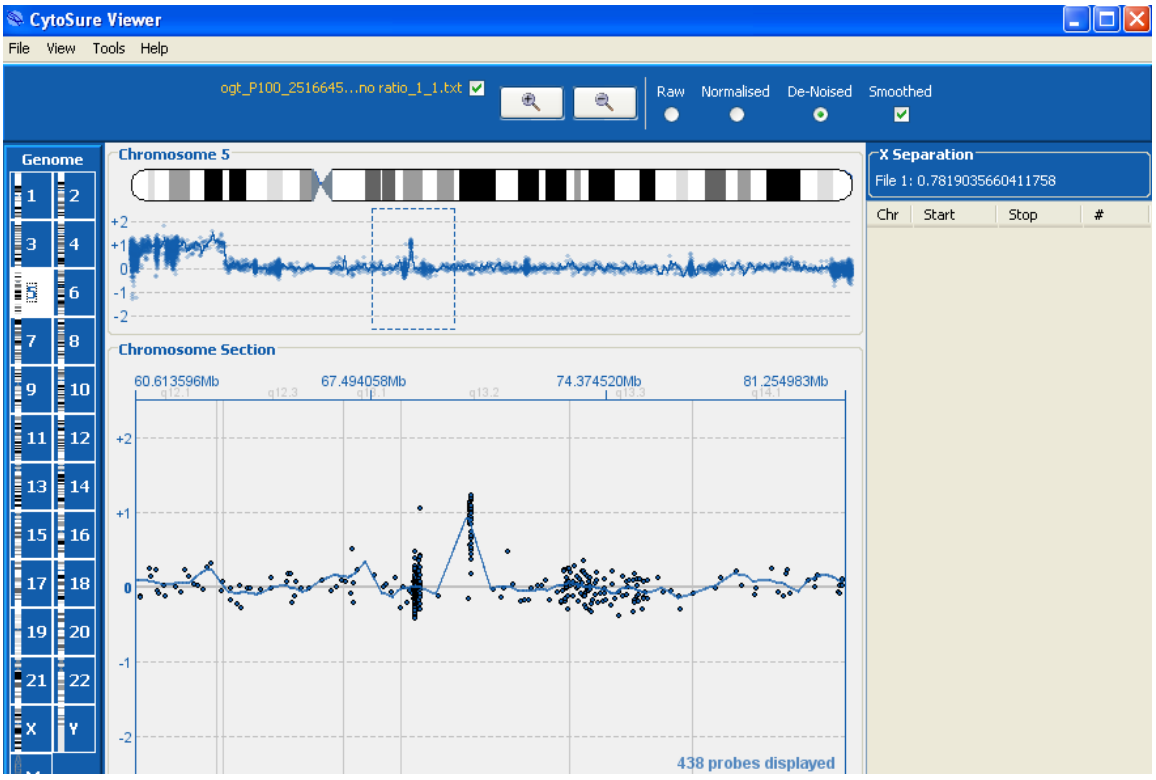


Figure 28 – Smoothed, normalised data.

4.7 Multiple datasets

It is possible to compare two or more datasets on the same plot by importing additional feature extracted files.

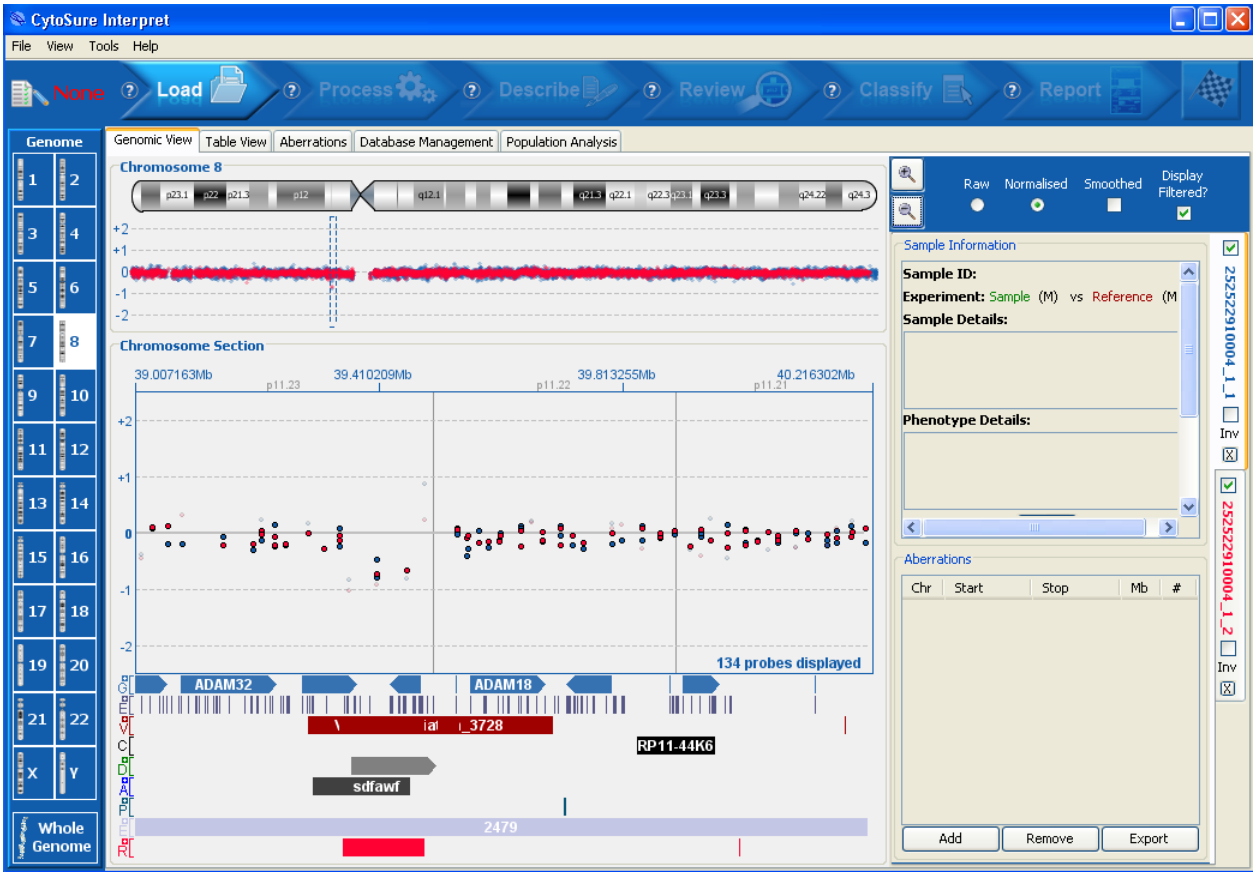


Figure 29 – Viewing two datasets. Note the two tabs at the right hand side of the screen

4.8 Customising the display

Switch to a vertical layout

To switch from a horizontal to a vertical layout select Tools-> Options -> Display and click Horizontal signal axis

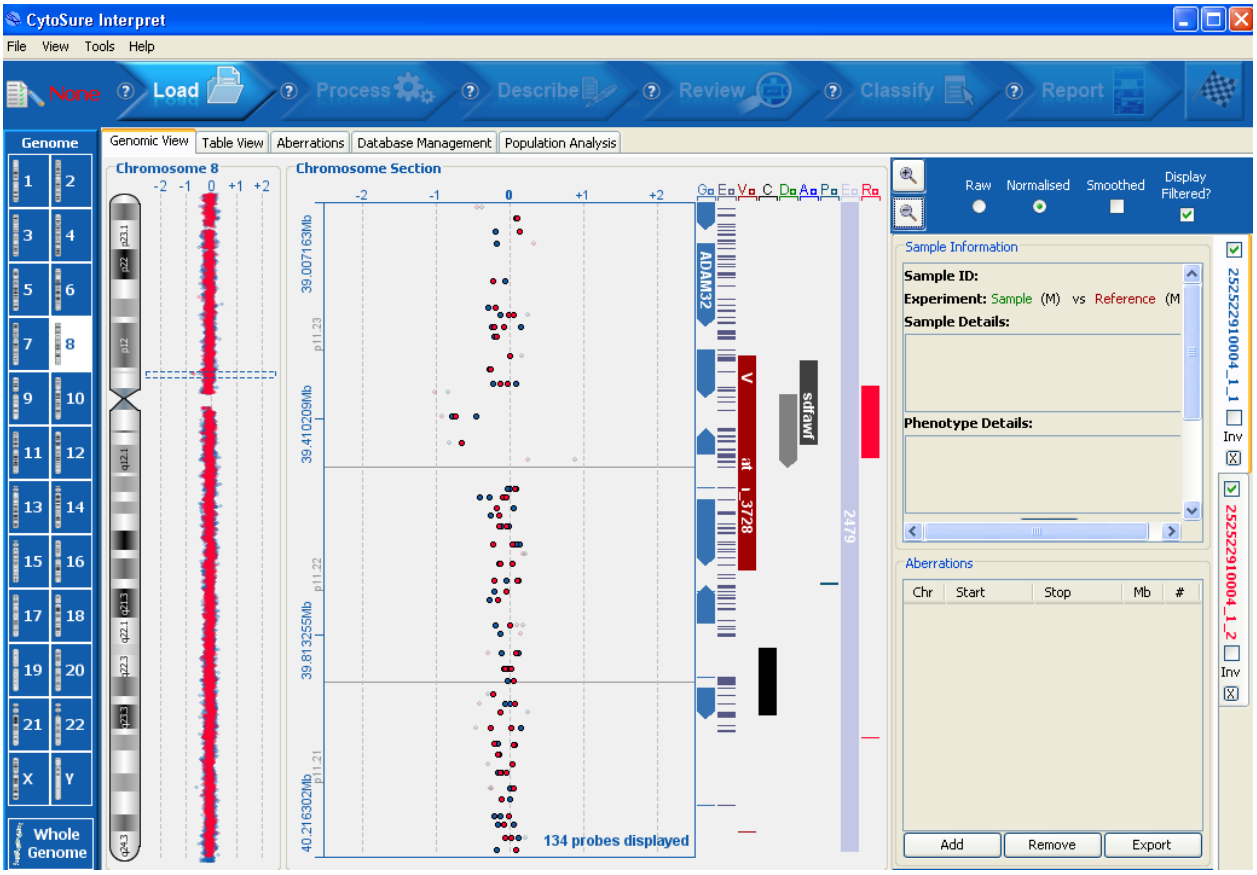


Figure 30 – Vertical display.

Colours

To switch the colours of the data points select **Tools** -> **Options.** Choose colour 1 or colour 2 and click **Edit.** Select colour required from the palette and click 'OK'.

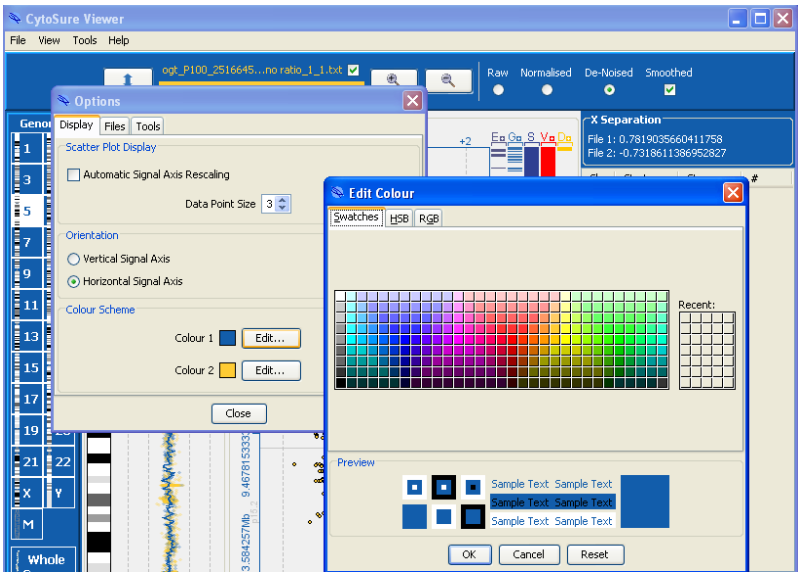


Figure 31 – Options for data point colours.

Altering the data points sizes

Select **Tools** -> **Options** and select the data point size required. Click 'Ok'.

Turning on/off axis rescaling

CytoSure automatically rescales the y axis, depending on the ratio values in that region. This can be confusing if comparing data points from different parts of the genome. The rescaling can be turned off by selecting **Tools** -> **Options** and checking the relevant radio button.

Axis options

There are several options in **View** that will rescale the scatter plot display, either increasing or reducing the y-axis or toggling up and down the axis. It might be necessary to first turnoff automatic rescaling (see section above).

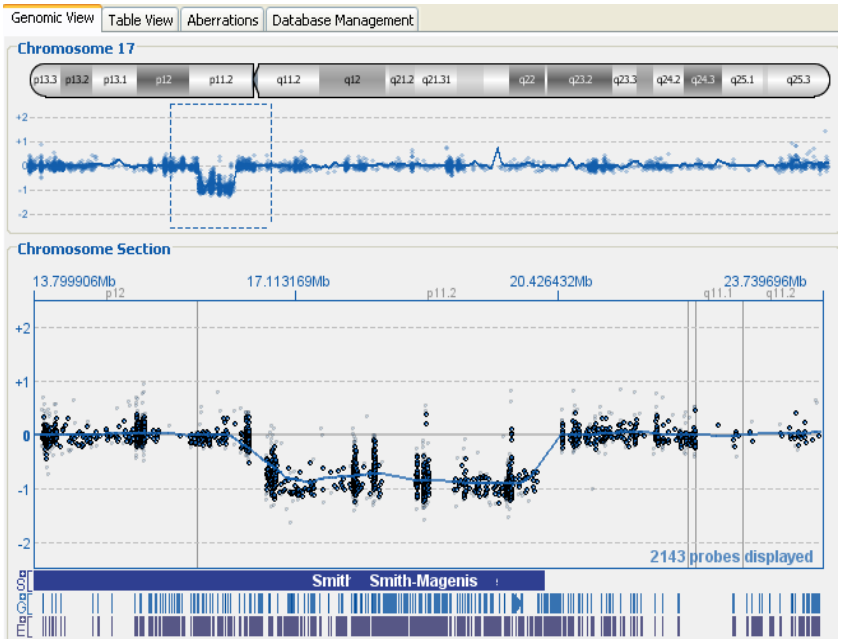


Figure 32 – Display with smoothing line added.

Changing ideogram view

To alter the shaded view of the chromosomes to a 2D view click Tools > Options > Karyotype Band Rendering.

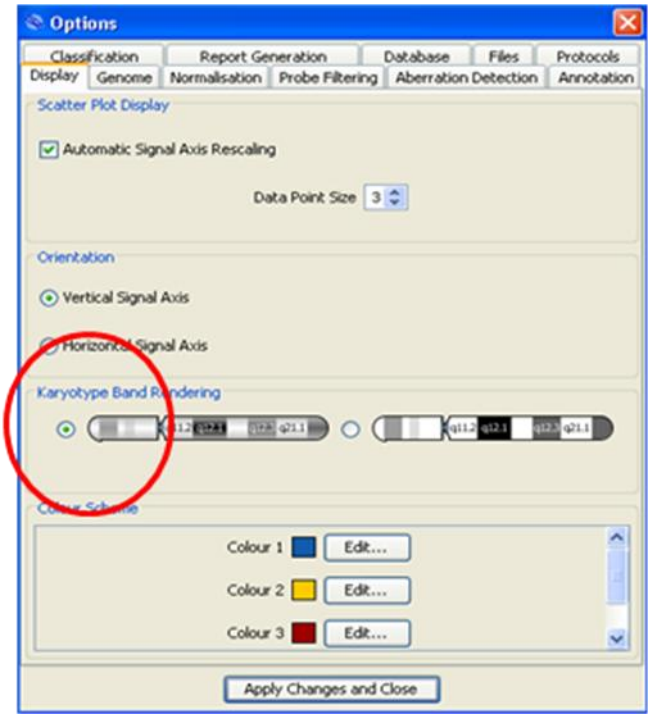


Figure 33 – Altering the Karyotype Band Rendering

.

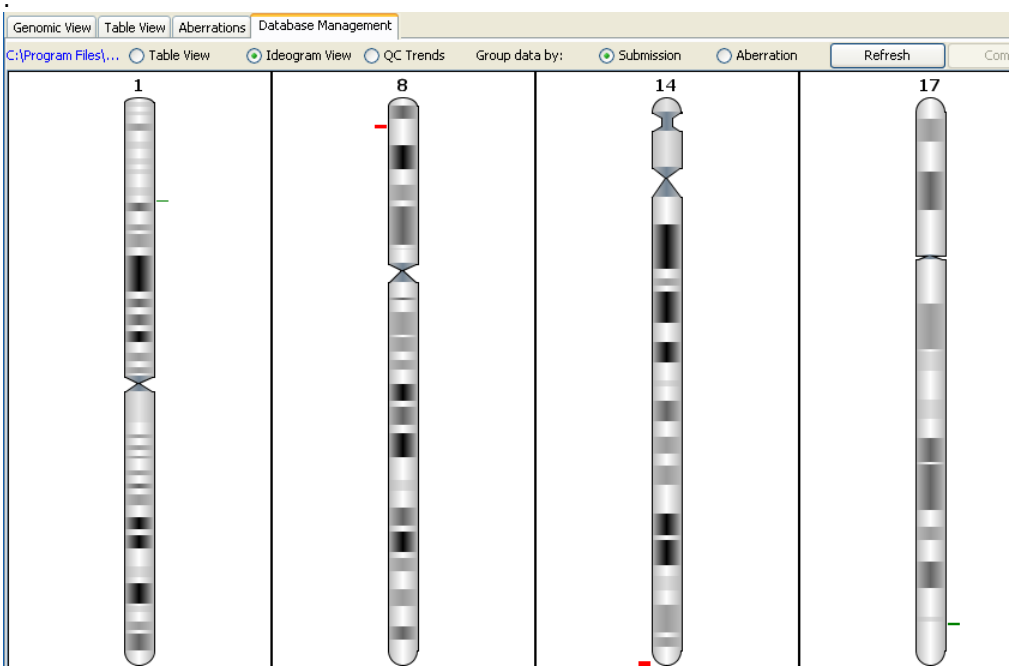


Figure 34 - ‘3D’ chromosomal ideogram.

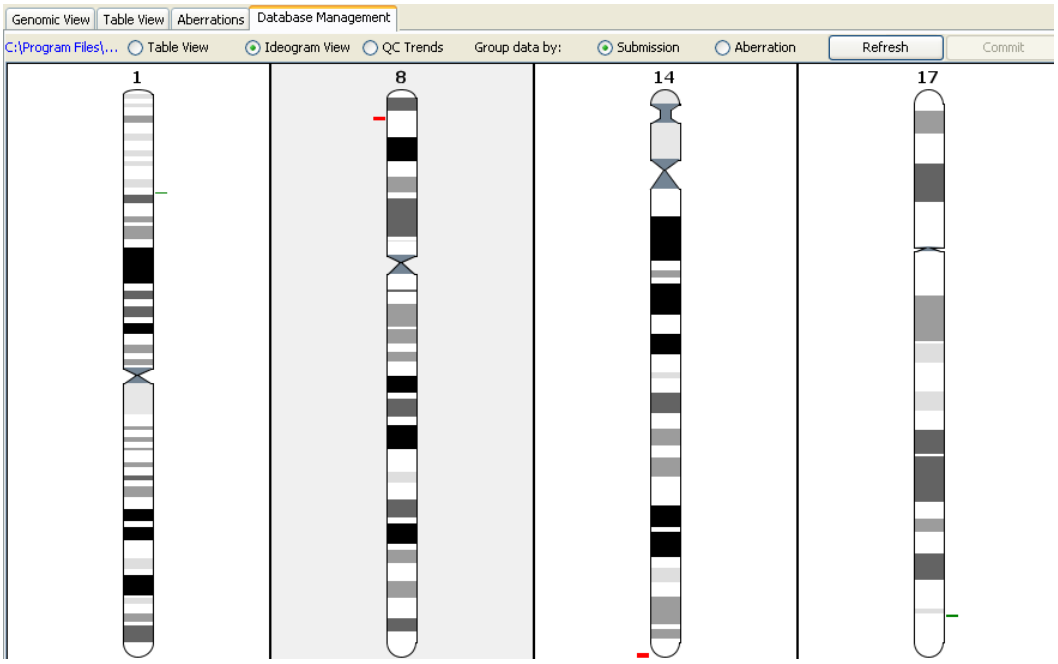
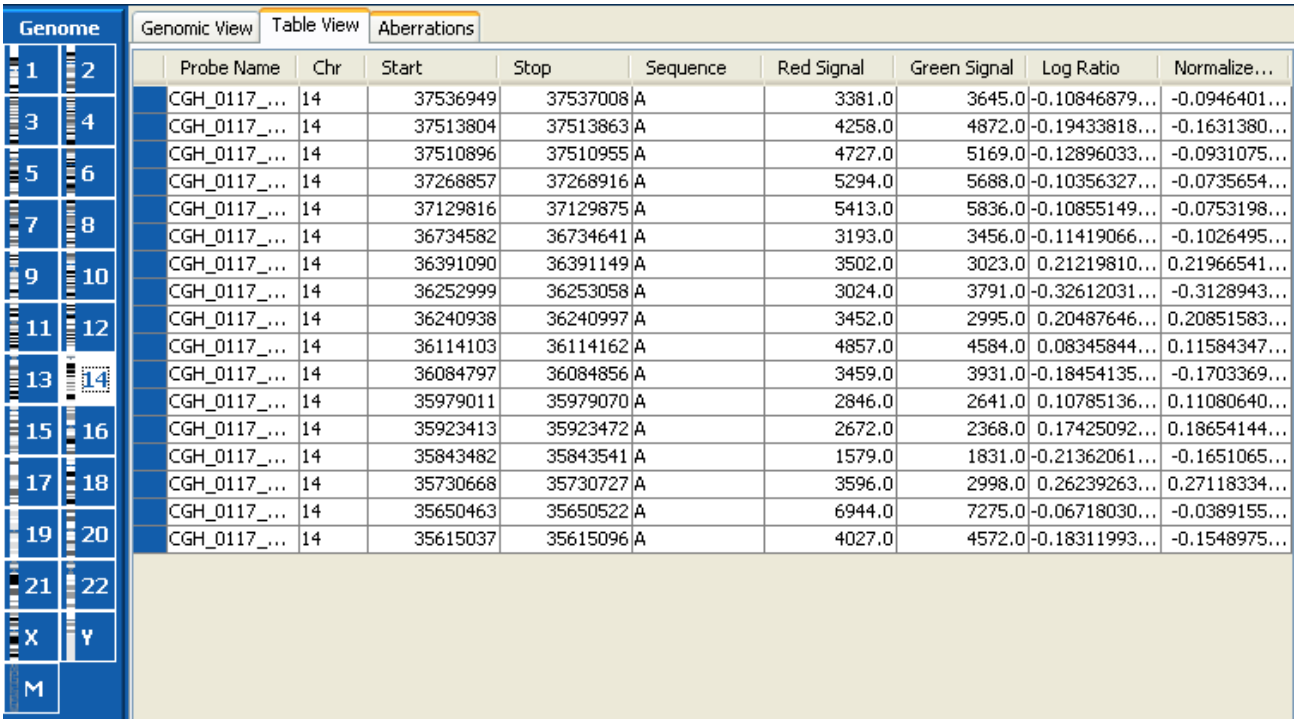


Figure 35 - ‘2D’ chromosomal ideogram.

4.9 Table View tab

To obtain additional information on displayed probes d, use table view. Navigate in Chromosome section view to display probes desired. Click on Table tab (E on figure 1). This reveals relevant probe information.



Probe Name	Chr	Start	Stop	Sequence	Red Signal	Green Signal	Log Ratio	Normalize...
CGH_0117_...	14	37536949	37537008	A	3381.0	3645.0	-0.10846879...	-0.0946401...
CGH_0117_...	14	37513804	37513863	A	4258.0	4872.0	-0.19433818...	-0.1631380...
CGH_0117_...	14	37510896	37510955	A	4727.0	5169.0	-0.12896033...	-0.0931075...
CGH_0117_...	14	37268857	37268916	A	5294.0	5688.0	-0.10356327...	-0.0735654...
CGH_0117_...	14	37129816	37129875	A	5413.0	5836.0	-0.10855149...	-0.0753198...
CGH_0117_...	14	36734582	36734641	A	3193.0	3456.0	-0.11419066...	-0.1026495...
CGH_0117_...	14	36391090	36391149	A	3502.0	3023.0	0.21219810...	0.21966541...
CGH_0117_...	14	36252999	36253058	A	3024.0	3791.0	-0.32612031...	-0.3128943...
CGH_0117_...	14	36240938	36240997	A	3452.0	2995.0	0.20487646...	0.20851583...
CGH_0117_...	14	36114103	36114162	A	4857.0	4584.0	0.08345844...	0.11584347...
CGH_0117_...	14	36084797	36084856	A	3459.0	3931.0	-0.18454135...	-0.1703369...
CGH_0117_...	14	35979011	35979070	A	2846.0	2641.0	0.10785136...	0.11080640...
CGH_0117_...	14	35923413	35923472	A	2672.0	2368.0	0.17425092...	0.18654144...
CGH_0117_...	14	35843482	35843541	A	1579.0	1831.0	-0.21362061...	-0.1651065...
CGH_0117_...	14	35730668	35730727	A	3596.0	2998.0	0.26239263...	0.27118334...
CGH_0117_...	14	35650463	35650522	A	6944.0	7275.0	-0.06718030...	-0.0389155...
CGH_0117_...	14	35615037	35615096	A	4027.0	4572.0	-0.18311993...	-0.1548975...

Figure 36 – Table view showing detailed probe information.

4.10 Aberration tab and ideogram view

To access the table of aberrations, click on the aberration tab (F on figure 1). This can be used to annotate those aberrations which have been generated either by automated aberration detection (section 5.1) or manually (section 5.2). See section 5.3 for more details.

To view aberrations on an ideogram click on the radio button (figure 37). The gains are indicated in green and the losses in red (see Fig 38).

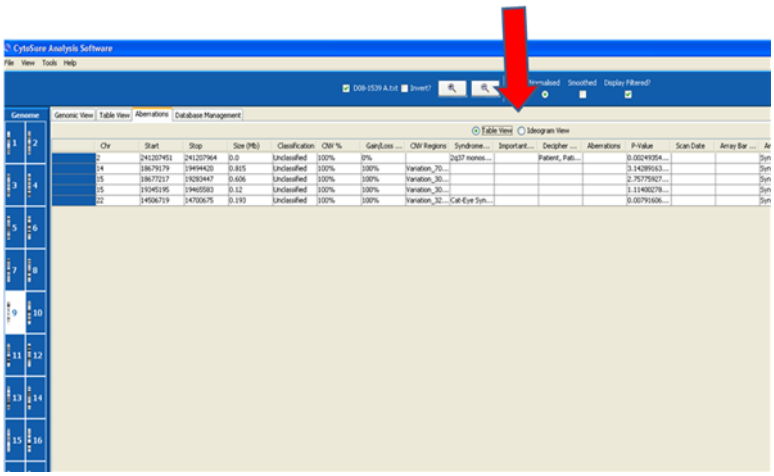


Figure 37 - The ideogram radio button.

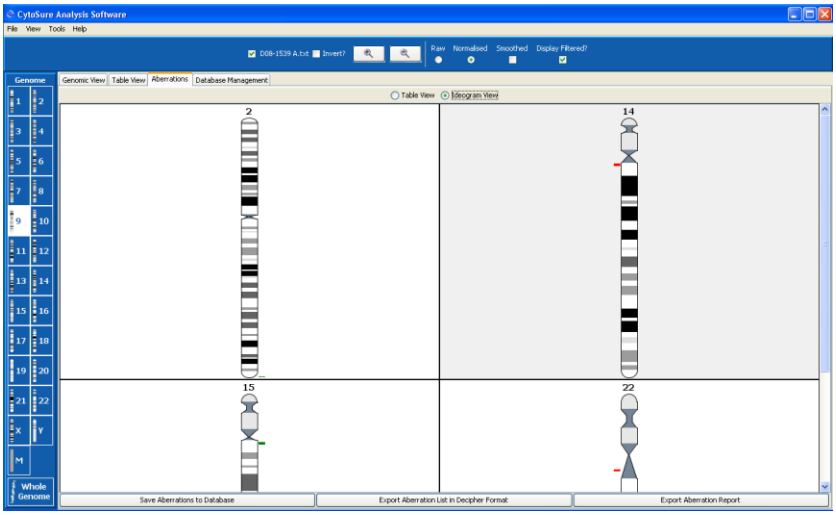


Figure 38 - The ideogram view

Click on an individual chromosome to view full screen. Re- click on the chromosome to return to the standard ideogram view.

4.11 Aneuploidy testing

This feature rapidly tests the presence of aneuploidies in the sample. In genomic view, click Tools > Aneuploidy Summary. This provides a rapid method for averaging all probes on each chromosome and plotting a box and whisker plot. The plot shows the following:

- position of the mean
- median (50th percentile)
- lower quartile (25th percentile)
- upper quartile (75th percentile)
- maximum value
- minimum value.

See figure 39 for details. Zoom into regions on the plot by dragging the mouse icon over the relevant region.

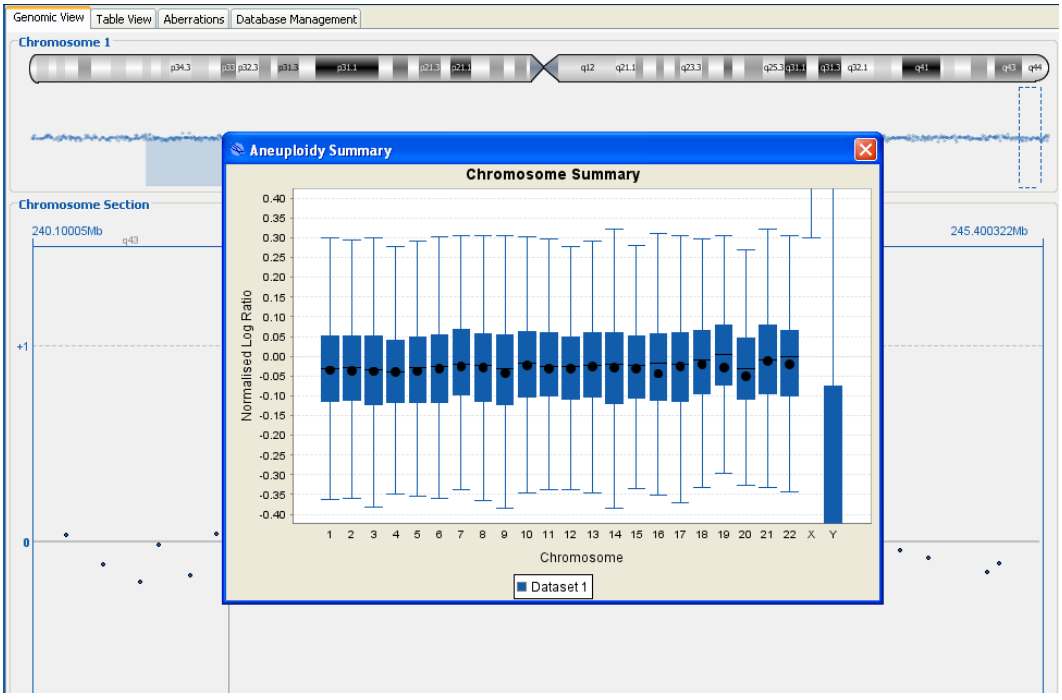


Figure 39 -Diagram of the Aneuploidy Summary screen.

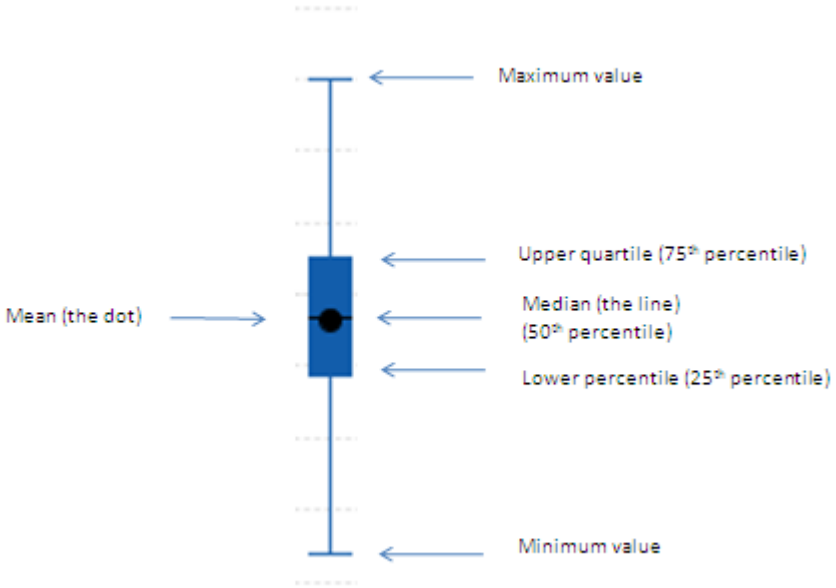


Figure 40 - Diagram explaining the different values on the box and whisker plot.

5. Adding aberrations to generate an aberration list, annotating, saving and exporting for printing

5.1 Automated aberration detection

The program uses a modified form of Circular Binary Segmentation (CBS) to divide the data into regions of the profile which have a similar copy number. These are known as segments (Venkatraman ES and Olshen AB *(*2007) *Bioinformatics* **23(6)** p657-663).

The relevant segments then need to be called as aberrant. The call is dependent on how many probes are in the segment and the $\log_2$ ratio value of the segment (known as the threshold value).

The method to run automated detection and aberration calling is as follows:

1. Click on Tools -> **Identify Aberrations**.
2. A dialog box displays the option to either run the CBS on filtered data (see section 4.5) or on raw data.

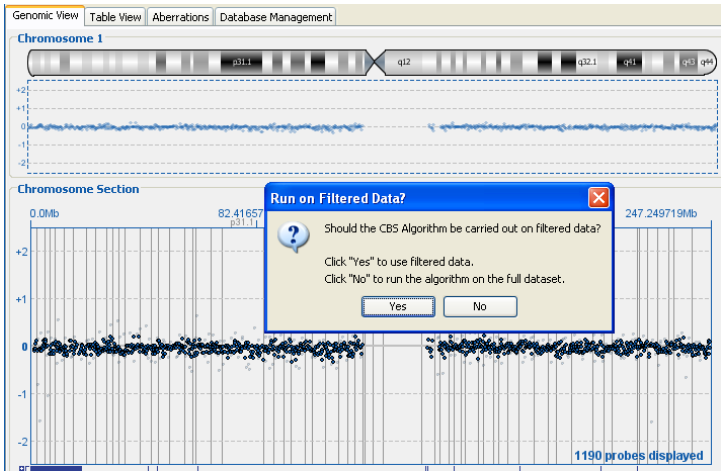


Figure 41 - Running the CBS on either filtered or raw data.

2. The CBS algorithm runs with the default settings. To change the settings click Tools -> Options -> Aberration Calling Options.
3. The CBS option box displays option settings for calling of the segments obtained by CBS

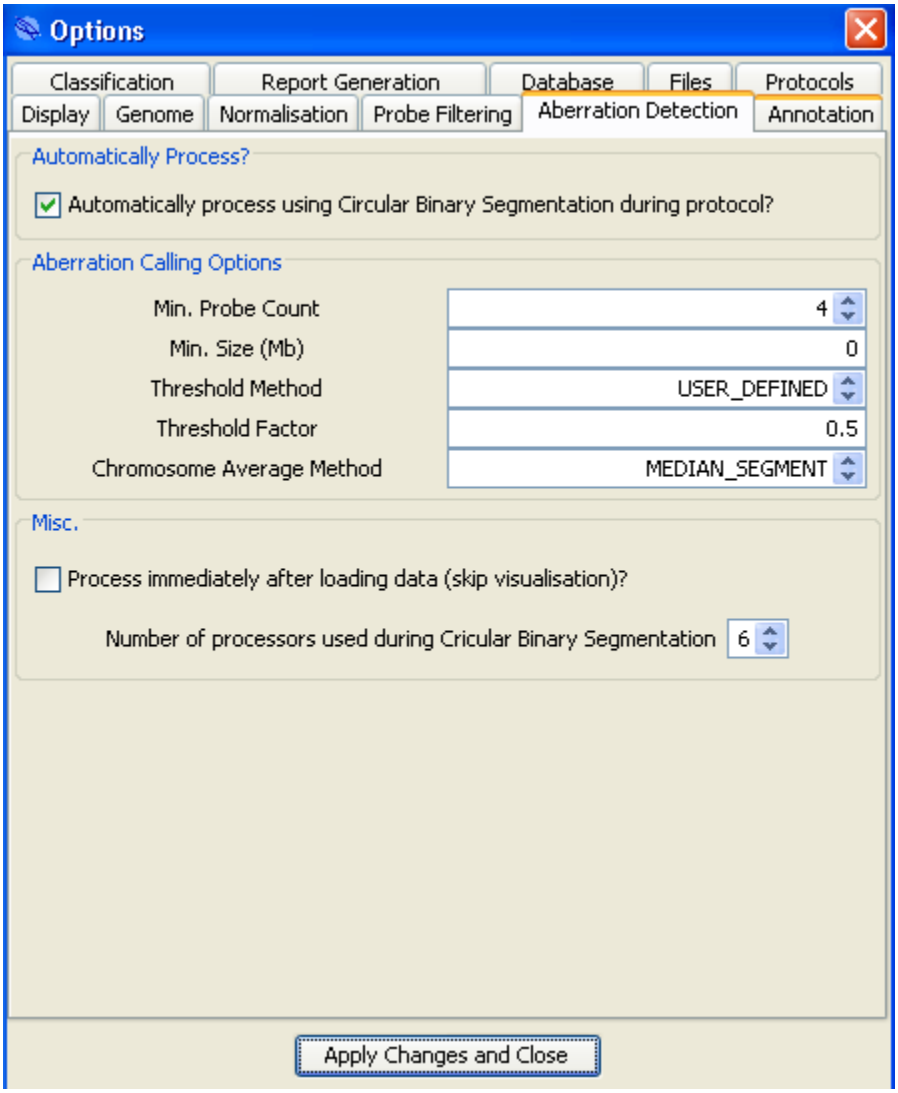

Options
Classification
Report Generation
Database
Files
Protocols
Display
Genome
Normalisation
Probe Filtering
Aberration Detection
Annotation
Automatically Process?
Automatically process using Circular Binary Segmentation during protocol?
Aberration Calling Options
Min. Probe Count
4
Min. Size (Mb)
0
Threshold Method
USER_DEFINED
Threshold Factor
0.5
Chromosome Average Method
MEDIAN_SEGMENT
Misc.
Process immediately after loading data (skip visualisation)?
Number of processors used during Cricular Binary Segmentation
6
Apply Changes and Close

Figure 42 – View of option settings for calling the aberrant segments.

Minimum Probe Count:
A value of 4 probes is suggested as the minimum number of probes in a segment that are required to make a call.

Threshold method and threshold factor:
This is the value that the algorithm uses to set the threshold value.

User definable – Type the value of the $\log_2$ ratio into the threshold factor box to set the threshold value that will be used to make a call. Typically a $\log_2$ ratio of 0.5 to 0.85 is used.

Standard deviation – the software calculates the standard deviation of the ratio ; the threshold value will be the threshold factor x standard deviation of the ratios. The preferred factor is 3 or 4.

Deviation Log Ratio (DLR) – the software calculates the DLR Spread of the ratio ;the threshold value will be the threshold factor x DLR of the ratios. This typically is a factor of 3 or 4.

X separation – The software uses the value from the X chromosome probes. This assumes that a sex mismatch experiment has been run. It may omit duplications, which usually need a lower threshold value compared to deletions.

Include X chromosome? Choosing False will exclude the X chromosome.

Include Y chromosome? Choosing False will exclude the Y chromosome probes. This might be used when Female sample and references are used.

Chromosome average method:
The method of averaging the segments at the baseline. This is used to set the threshold value.

Median segment uses the Median value of all segments in the chromosome as the baseline, and is the preferred option.

Mean segment uses the Mean value of all segments in the chromosome as the baseline. This could be affected by large aberrant regions in the chromosome.

Zero will set the baseline at zero whatever the data is.

Chromosome Mean log probes –Uses the mean value of all the probes as the baseline

Chromosome Median log probes – Uses the median value of all probes as the baseline

4. Click done.
5. The CBS will start running, segmenting the data.
6. The number of threads can be increased. The more threads used the faster the processing, however alternative processes on the computer will be slower. A quad computer using three threads should take approximately 4 - 5 minutes to process data from a Syndrome Plus 2x105k array.
7. When the CBS is complete horizontal lines will be drawn representing the segments.
8. Aberrations called are represented by shaded areas. Aberrations will be placed in the Aberration window where they can be edited if required.
9. It is recommended that once the CBS has run, the file is saved using **File -> Save.** The CBS results are saved.
10. If the threshold values need to bealtered, there is no need to rerun the CBS. Select **Tools -> Options.** Selection of the Tools tab and the CBS option, changes the threshold values (see figure 35).

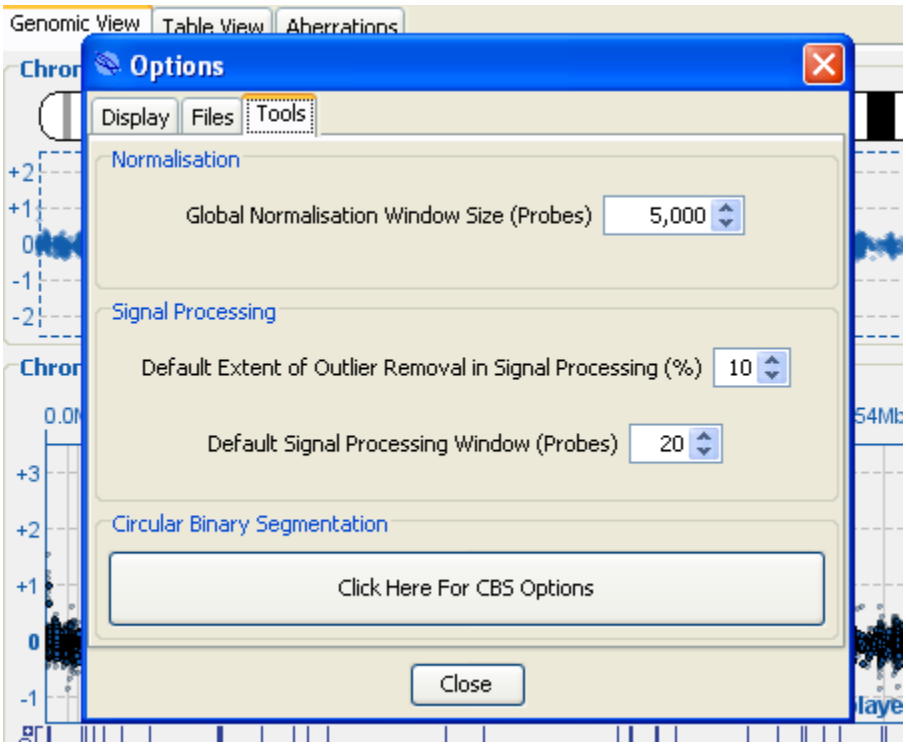


Figure 43 - Changing the CBS threshold values

.

5.2 Batching automated aberration detection

The software enables multiple files to be segmented using the CBS in a batch process. This means that the software can be left unattended to process the data.

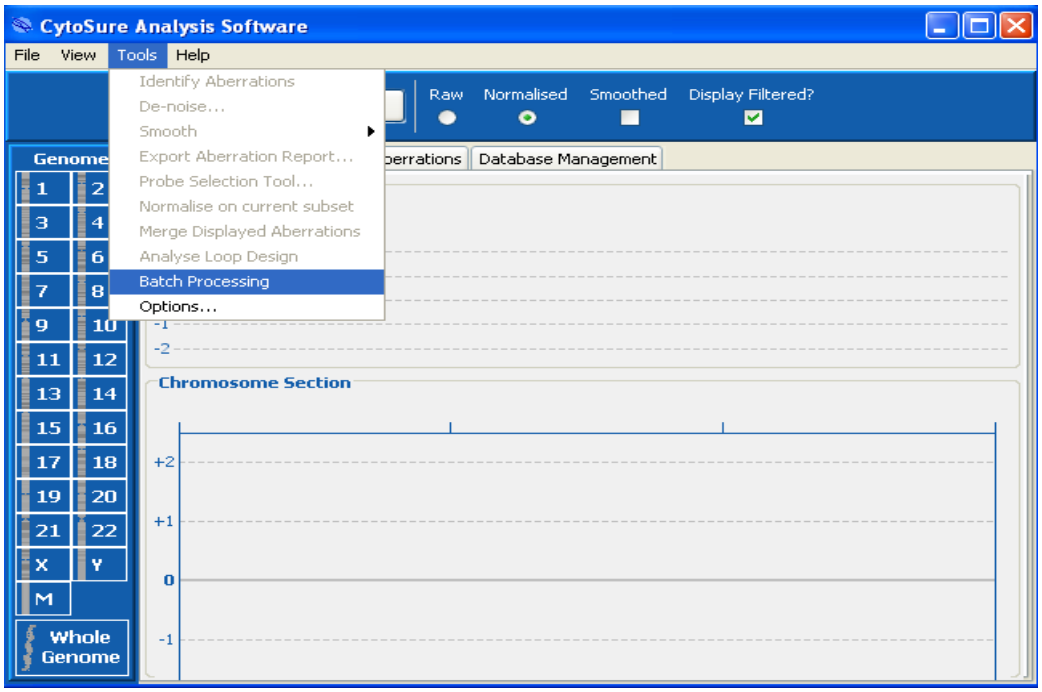


Figure 44: - Starting batch processing.

1. In Tools, click on Batch processing (see figure 44).

2. A dialog box opens (figure 45).Complete the following:

- Select a folder to store the processed data
- Add the file to the Files to be processed area. If the files are in GenePix format, then in the file type drop down menu ‘All files’ needs to be selected to display the .GPR files.
- Select number of threads to be used. The more threads used the faster the processing, however alternative processes on the computer will be slower. A quad computer using three threads should take approximately 5 minutes to process data from a Syndrome Plus 2x105k array.
- Select whether the data should be normalised or not
- Select whether the data used should be filtered or not

- Click Process

3. The processing will begin and the data saved to the appropriate folder.

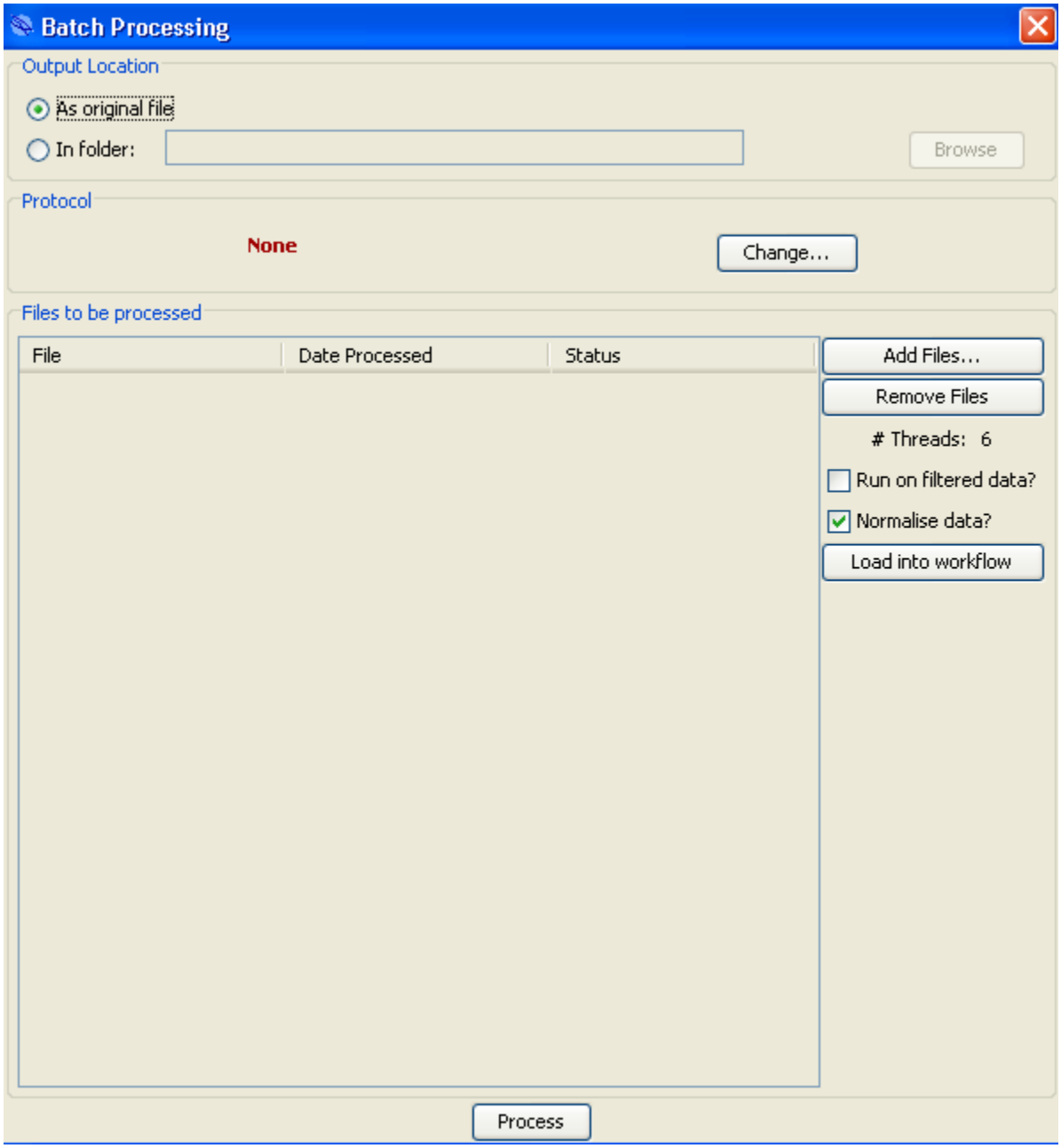


Figure 45 - CBS batch processing box.

5.3 Manual aberration detection

Manual aberration detection occurs when CBS processing is not selected.

1. Locate the aberration in either the chromosome view or whole genome view.

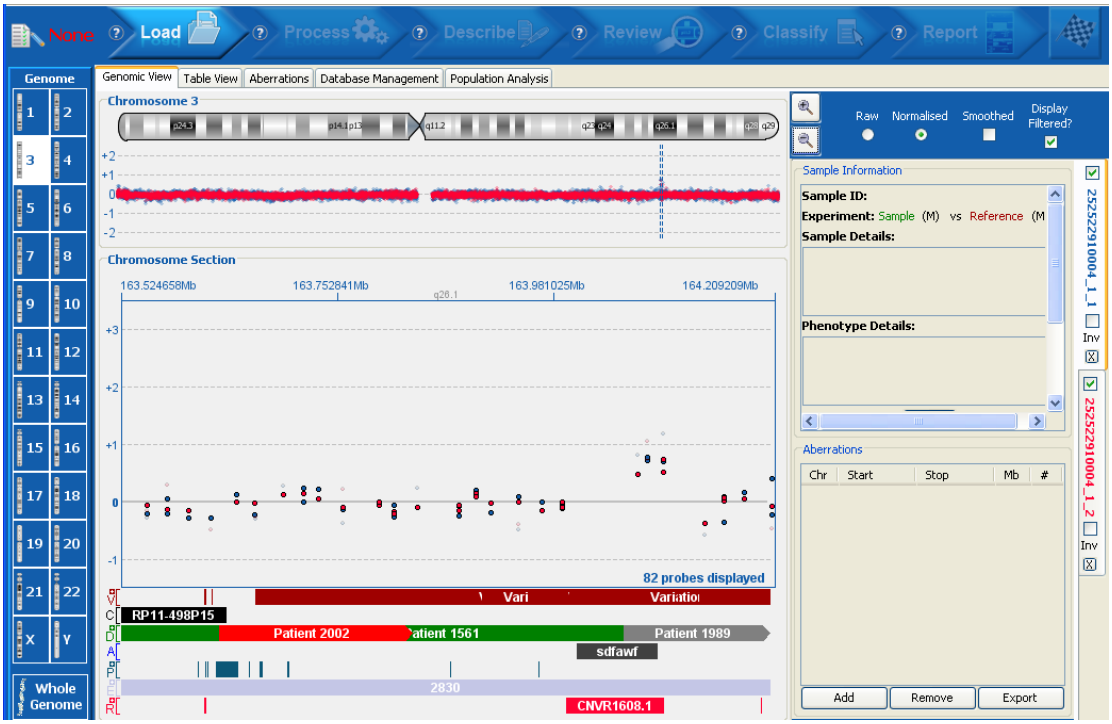


Figure 46 – Locate the aberration.

2. Zoom in. **All probes shown on the screen should be within the aberration**.
   There should be no flanking probes visible; otherwise these will be included in the export.

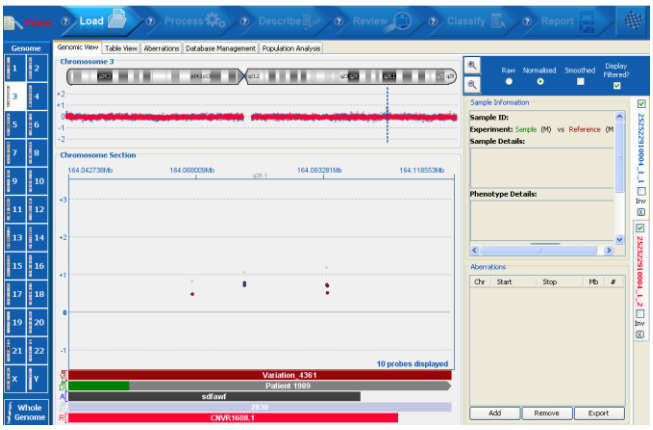


Figure 47 – Expand so only the data points in the aberration are visible on the screen.

3. Click Add. Repeat for other aberrations in the chromosome. **It is important to include all aberrations on the same chromosome (see discussion below)**

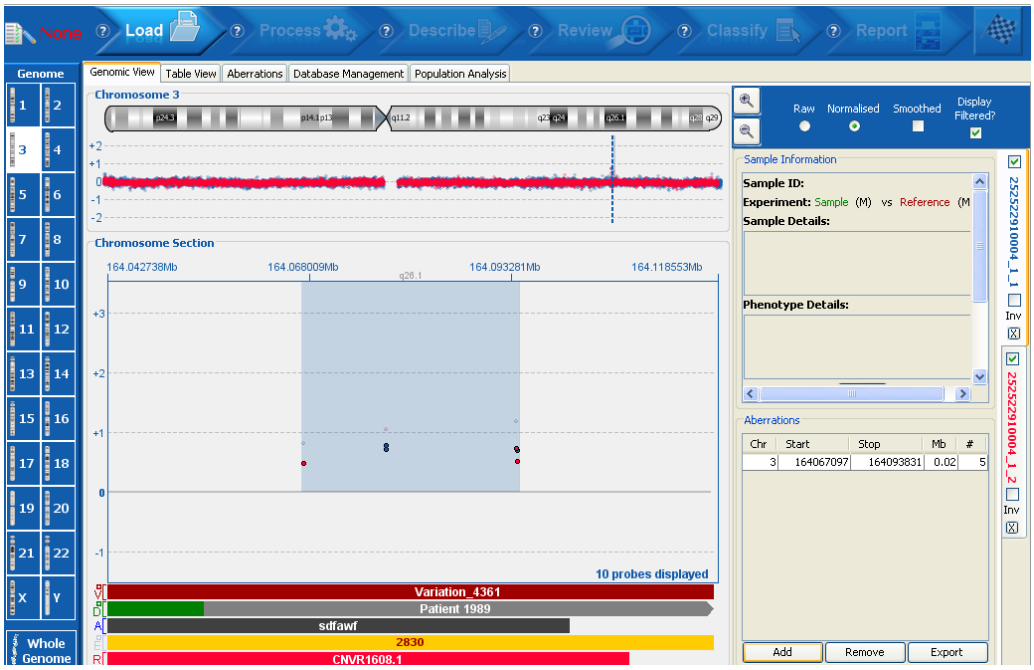


Figure 48 – Adding an aberration to the list

5.4 Link to SUSPECTS

A link to SUSPECTS has been added in the genomic view. SUSPECTS is a site that uses publication information and **theoretical** calculations, such as gene expression data, to highlight genes in an aberration that may be disease causing. For more details visit the SUSPECTS website. In genomic view, with an aberration present, right click on the aberration and select SUSPECTS

Figure 49: Selecting the link to SUSPECTS

A box will appear with the option of typing in the phenotype. SUSPECTS can use a training set of genes if completed, however it is not necessary and can be left blank.

5.5 Editing and Annotating the Aberration list and saving

5.5.1 Merging aberrations

The Segmentation may split an aberration into multiple segments (see figure 50). These will be present as multiple entries in the aberration table..

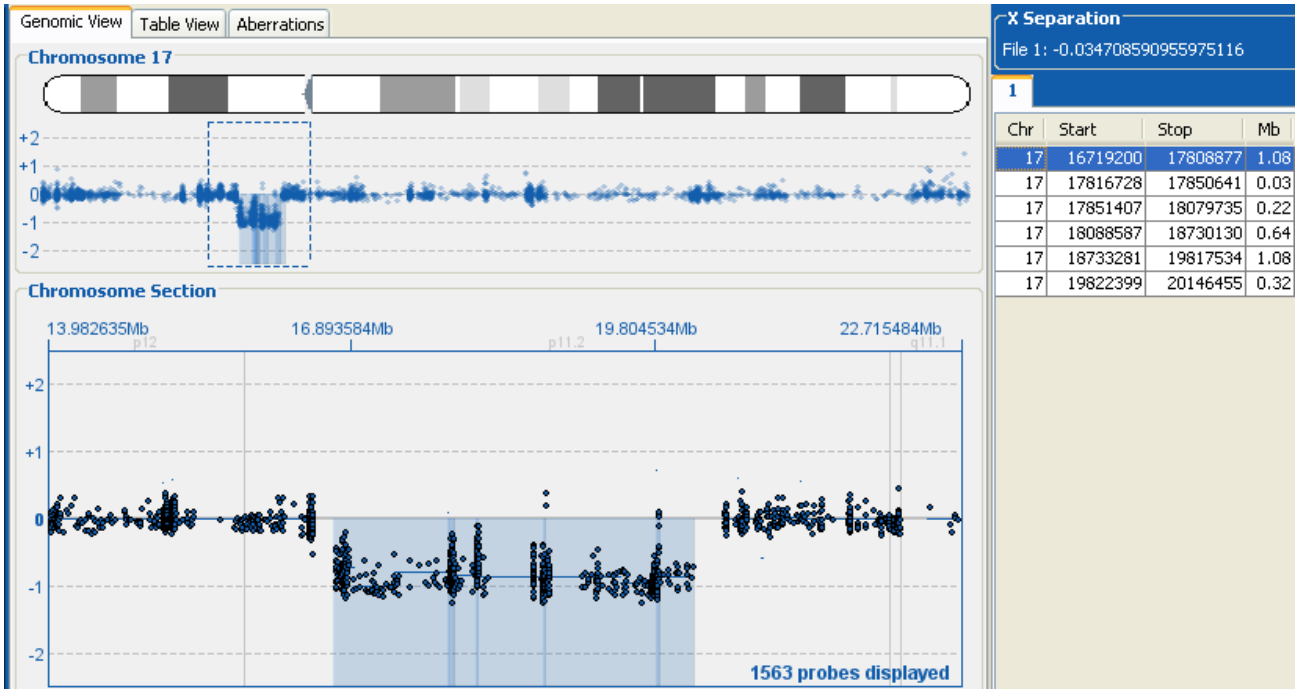


Figure 50 – An aberration spilt into multiple segments.

It is possible to merge these segments using the following procedure:

1. Navigate in the data view to reveal the aberrations to be merged
2. Click Tools-→ Merge Displayed Aberrations (see figure 51)
3. The aberrations will be presented as a single aberration on the screen and in the aberration table

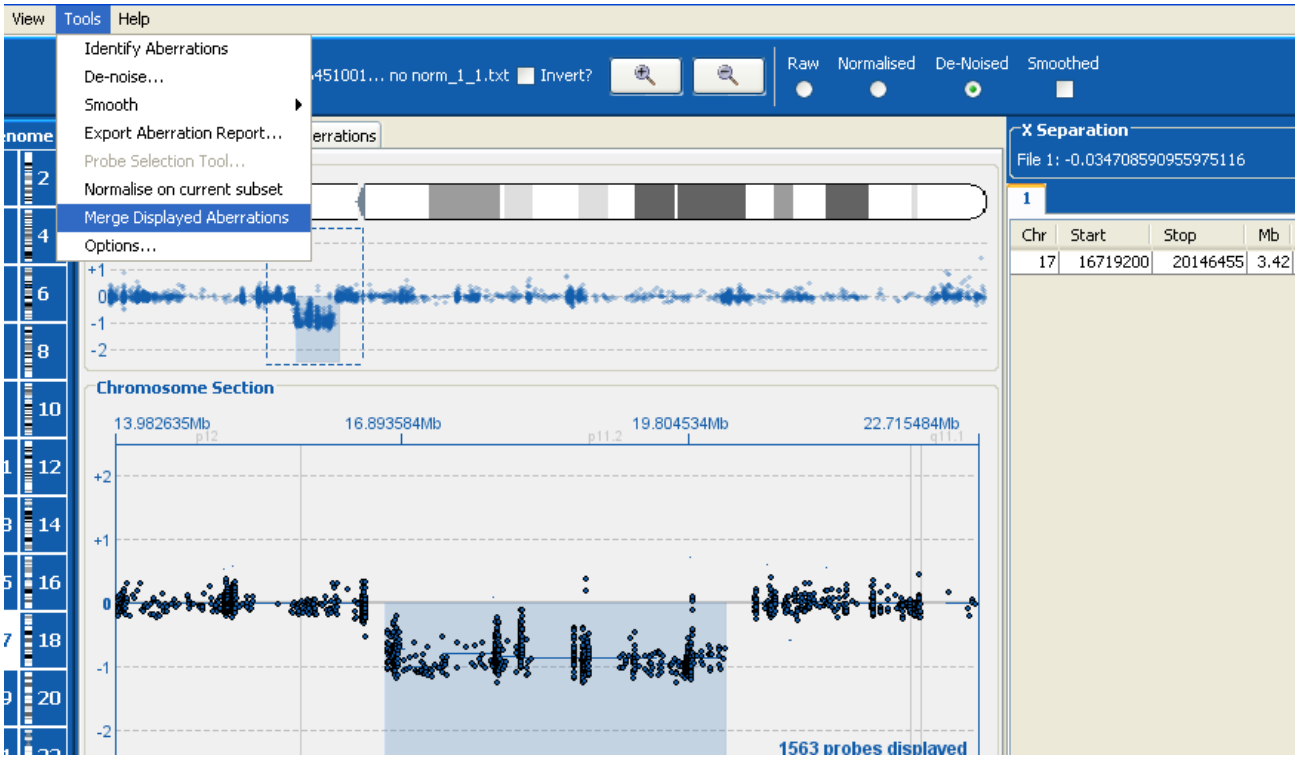


Figure 51 - Merging the aberrations.

To undo merging, place the mouse over the shaded region and right click. Select dissolve.

5.5.2 The Aberration tab: Editing and annotating the aberration list

Altering the sample details after File>Import

The sample details annotation can be altered after the File has been imported using File -> Properties (see figure 52) or using the Edit button in the Sample details panel. If using File -> properties, select 'Click here to edit sample details' button to change the sample annotation. Click on the genome web view tab to refresh the aberration list screen, if required.

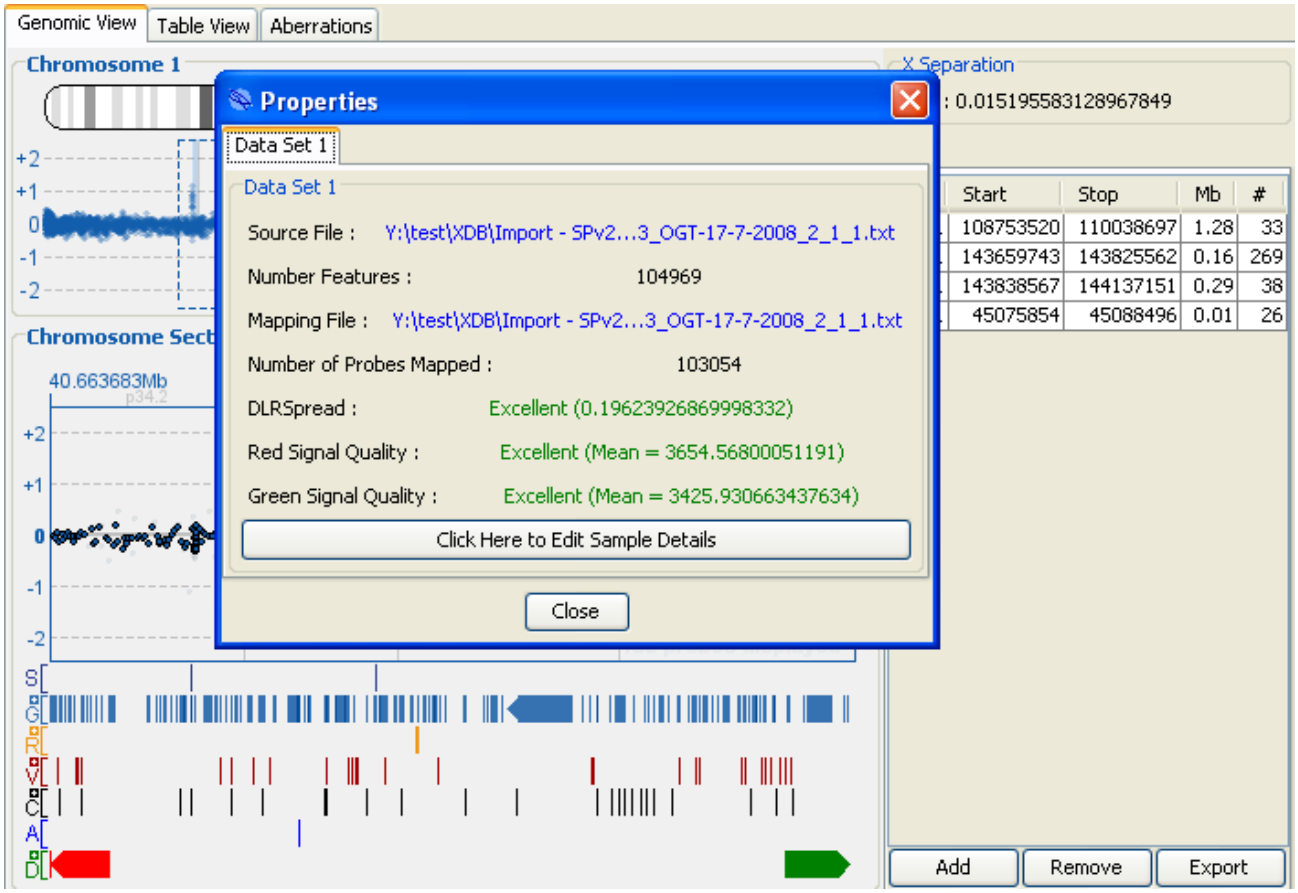


Figure 52 - Altering the sample details after File import.

The annotation list view

Once an aberration list has been generated, either manually or using automated detection, the list can be examined in more detail by clicking on the aberration tab (F in figure 1).

The aberration view shows a detailed list of the aberrations. Each column can be sorted by clicking on the header of the column.

The annotation provided is shown below:

Annotation	Description
Chr	
Start	Start of aberration. This is automatically entered
Stop	End of aberration. This is automatically entered

Size	Size of aberration. This is automatically entered
Scan Date	This is automatically entered
Array design	Depending on the array used, this is automatically entered
Gain/Loss	This is automatically entered
Syndrome	If the aberration overlaps any Syndrome annotation, then this is highlighted automatically.
Cytogenetic Location	Highlights the Cytogenetic banding location of the annotation
CNV	Reports % of the aberration that is annotated as a CNV. Does not take into account the DGV classification of whether the CNV is a gain or loss.
Gain/Loss CNV	Reports % of the aberration that is annotated as a CNV taking into account the DGV classification of whether the CNV is a gain or a loss.
DB CNV%	Reports % of the aberration that is annotated as a CNV in the internal database. Does not take into account the database classification of whether the CNV is a gain or loss.
DB Gain/Loss CNV%	Reports % of the aberration that is annotated as a CNV in the internal database taking into account the database classification of whether the CNV is a gain or loss.
CNV regions	This reports the CNV regions covered by the aberration. It does take into account that the CNV in the DGV database is a gain or a loss. If DGV does not report if it the CNV is a gain or a loss, then the CNV is reported.
Aberrations	Reports if the aberration overlaps with aberrations in the database
Decipher	Reports if the aberration overlaps with the decipher database (go to https://decipher.sanger.ac.uk/) for updated Decipher information
Important genes	Highlights human disease genes. Note that this list of human disease genes is not exhaustive
Detection	Displays whether the detection was by CBS (automatic) or manual
Automation Level	If for example an aberration was merged
P-values	Automatically generated. See section 5.8 for a discussion of the P-value
Comments	User editable. See section below.
Image Location	
Classification	The user can classify the aberration as follows: Unknown de novo - parental origin of rearrangement undefined de novo - arising on maternal chromosome de novo - arising on paternal chromosome familial (pat) - inherited from normal father familial (mat) - inherited from normal mother familial (mat) - inherited from mother with similar phenotype to child CNV (seen in normal individuals) CNV:parents not analysed CNV:parents analysed (seen in parents) CNV:parents analysed (not seen in parents) Familial parental origin unknown

	familial - inherited from normal mother and father Unclassified These terms can also be customised by the user. Please see section on customisation of classification terms in section 5.5.3.

On classification of the aberration, the row will change colour according to the classification assigned, as follows:

- Pink – CNV
- Grey – unclassified
- Purple – potential pathogenic aberration

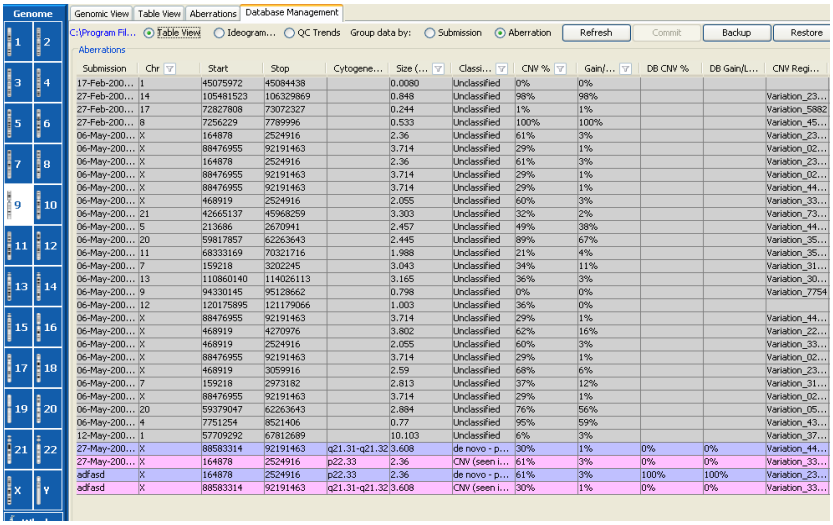


Figure 53: Colour-coding according to classification.

Editing, annotating and classifying the aberration list in Aberration view

To edit an aberration entry, right click on the desired aberration and choose **View Aberration** .This displays the aberration in genomic view.
To annotate the aberrations (see figure 54) select **Edit Aberration Details**. The aberration is classified according to the categories described above. Once the changes have been made, click Apply changes and Close. It is possible to set your own classification terms. Please see customisation of classification terms below.

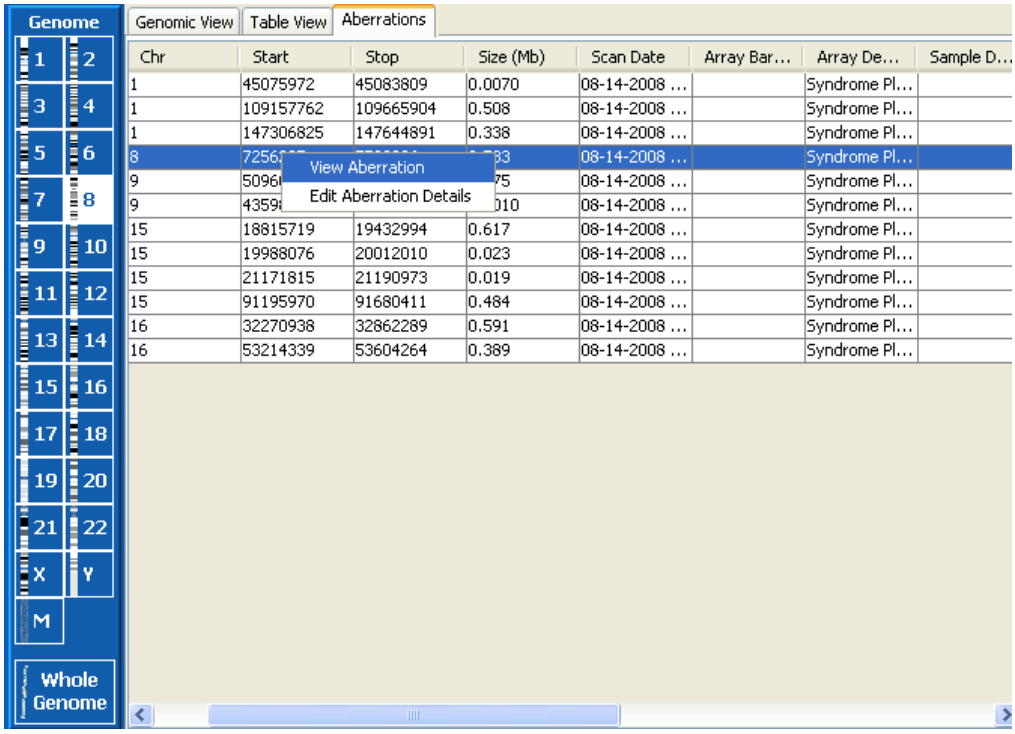


Figure 54 - Right clicking on the aberration in the Aberration view.

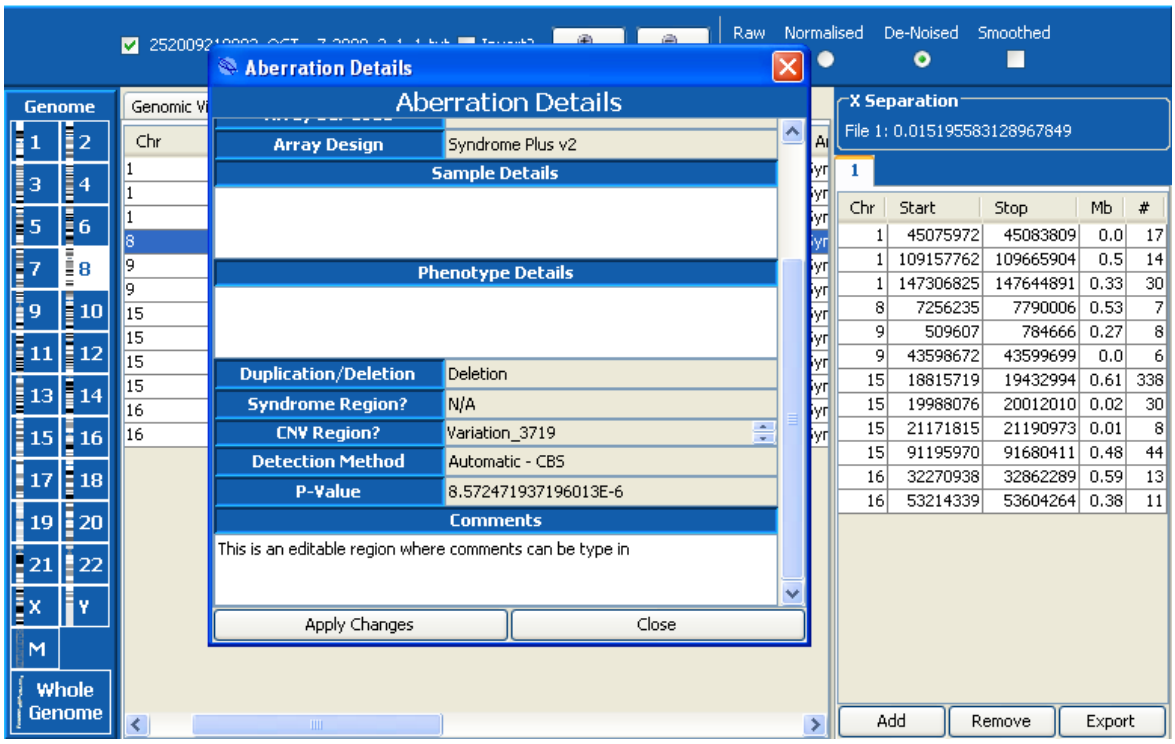


Figure 55 - Dialogue box to add comments on a particular aberration.

Editing and annotating the aberration list in Genomic view

In Genomic view, navigate to an aberration and right click (see figure 55) to reveal the following options:

Edit: Annotate the aberration (as described in 5.4.2)

Delete: Delete the aberration from the aberration list

Ensembl: View the aberration in Ensembl

UCSC: View the aberration in UCSC

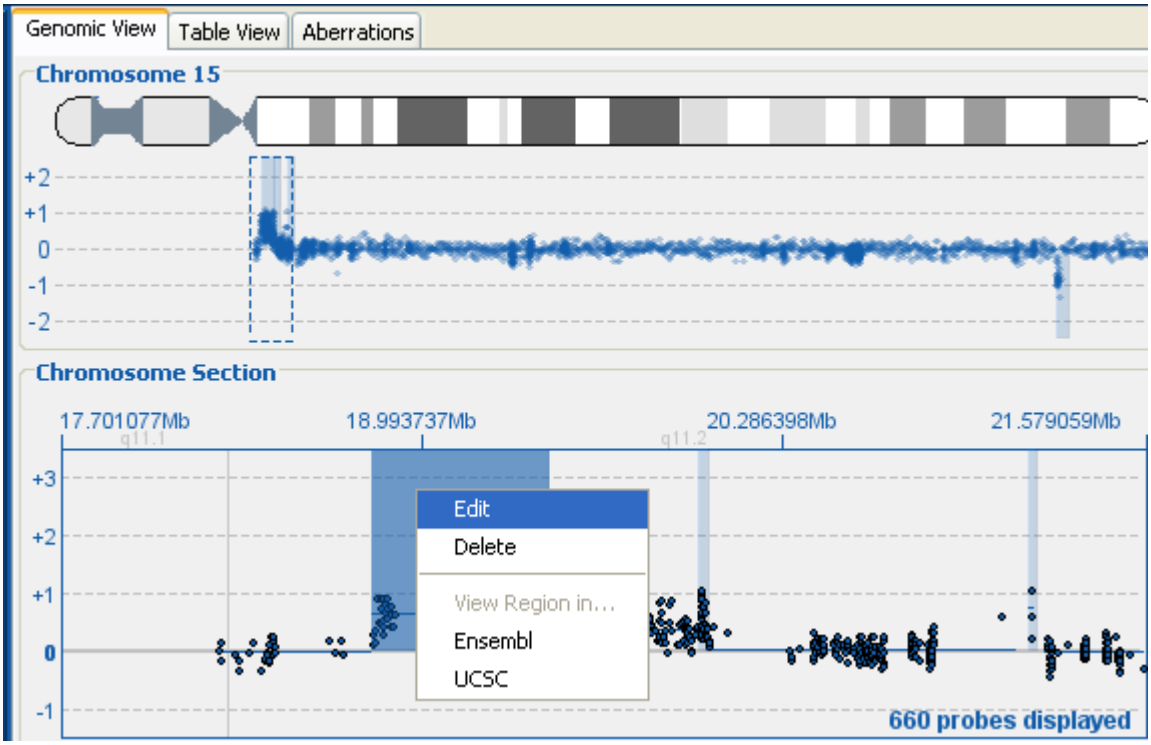


Figure 56 – Editing an aberration in Genomic view.

5.5.3 Customising the classification terms

Customising classification terms allows the user to define their own classification terms. Select Tools > Options > Classification

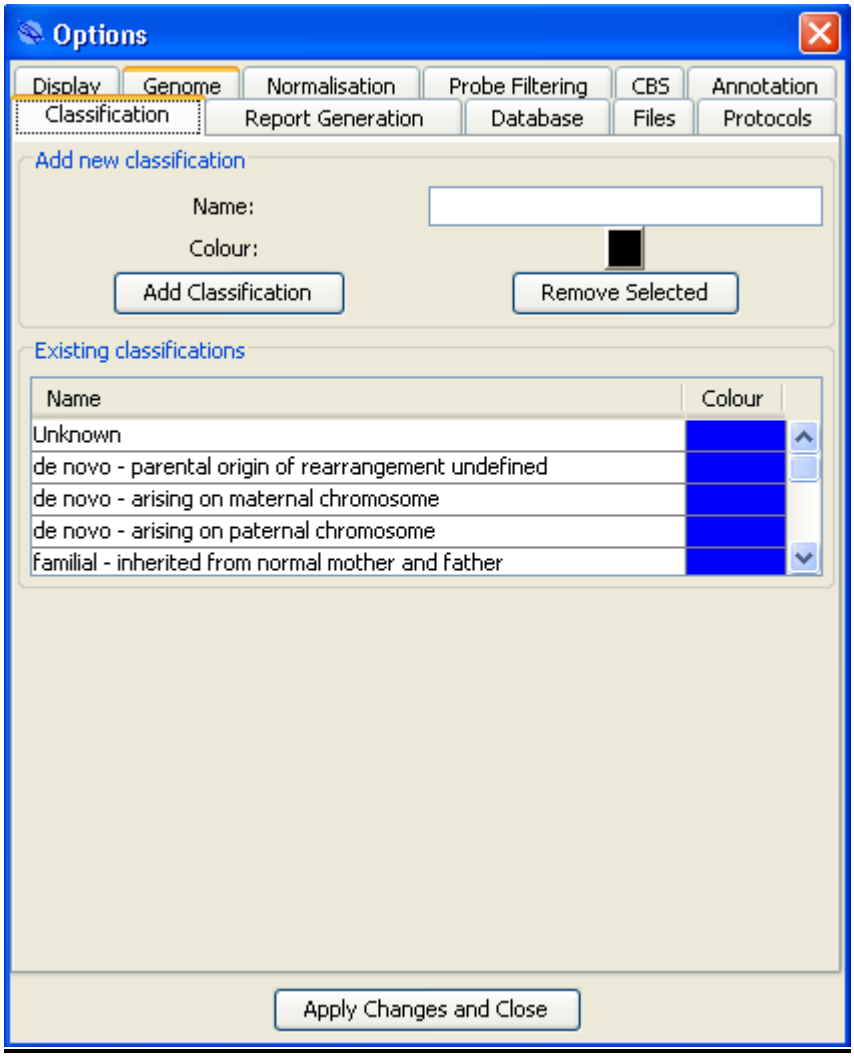


Figure 57 – Adding custom classification

The user can type in the name of the classification in the relevant box and then click on the colour box to change the colour. On clicking ‘Add Classification’ a new classification term will appear. This new classification term can then be used in the Aberration table.

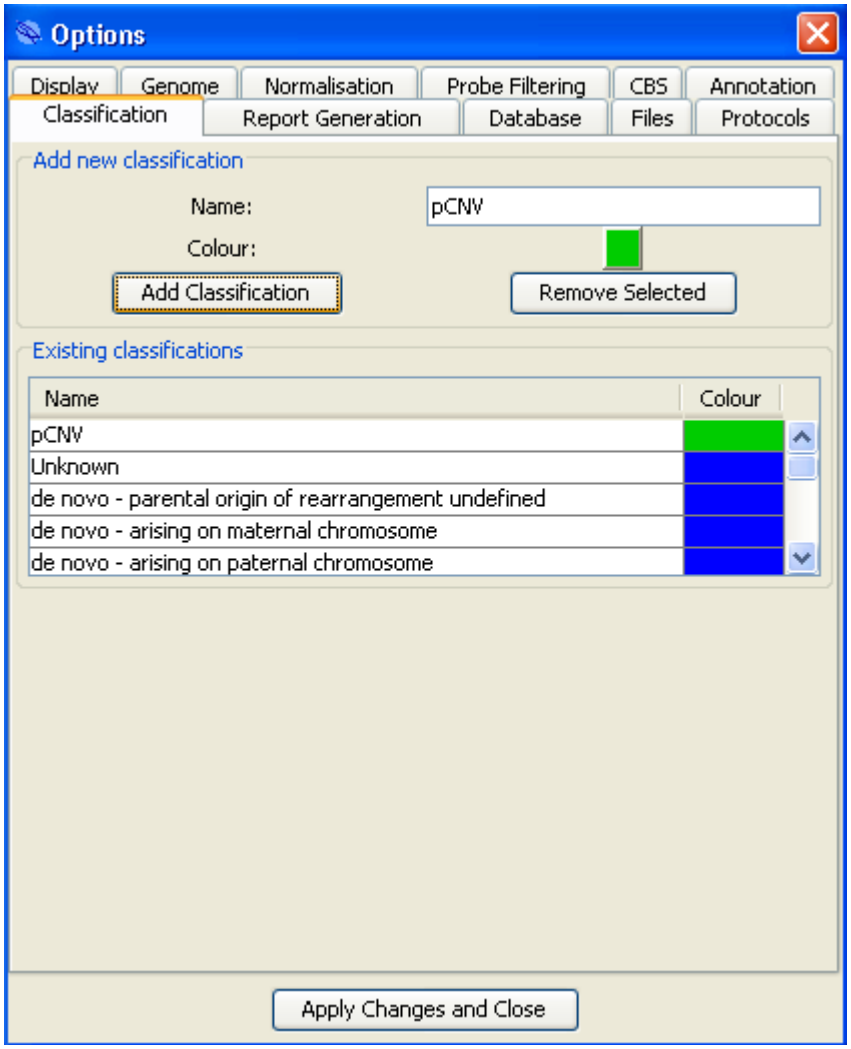


Figure 58 – Adding custom classification

The results of a custom classification that can be used in the aberration table are shown below

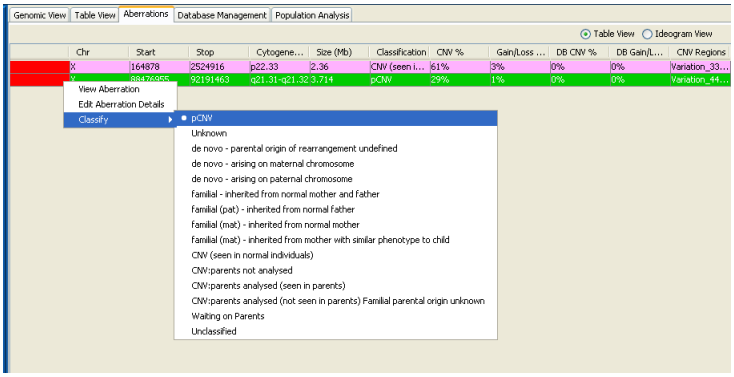


Figure 59 – new classification added

5.5.4 Saving the file

To save the Circular Binary Segmentation, the aberrations and the annotation use File -> Save.

Saving of multiple datasets could be limited by the memory capacity of the computer. To save 3 datasets it is recommended that the computer has a memory capacity of 3 to 4 GB.

5.5.5 Saving to the database

To save the annotated aberrations and images to the database, click on **Save Aberrations to Database** at the bottom of the screen (see figure 60 below).

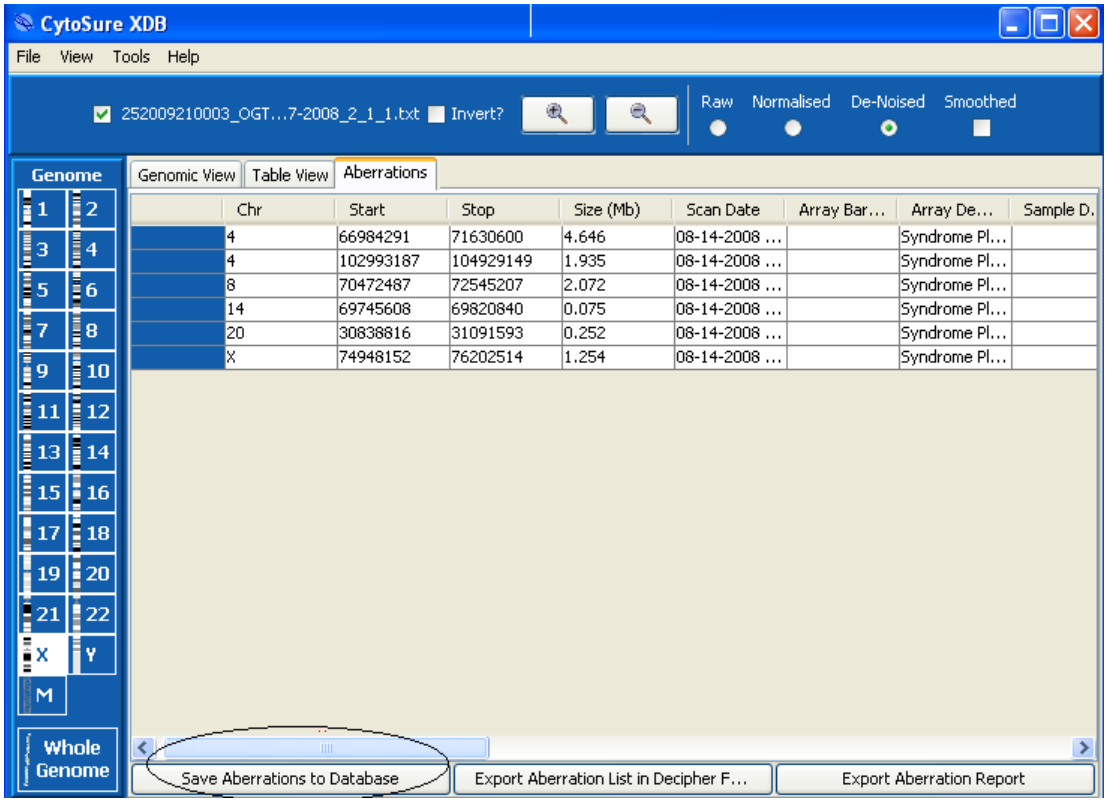


Figure 60 – Saving aberration annotation and images to the Database.

Saving data to the database has two effects:

1. The aberrations will be displayed in the aberration track within the Genomic View tab.
2. A new entry will appear in the Database Management tab.

5.6 Exporting aberration results for printing

1. Following generation of an aberration list,
either manually, using automated detection or a combination of both, click on the **Export** button. This can be done either in the Genomic view or the Aberration view (**Export Aberration list)**. This will generate a report.

2. The report can be generated as an html file, pdf file or a txt file with a separate folder for the images.

3. Open the report file. With the html file can be opened directly in Microsoft (MS) Word. Alternatively, 'Select all' 'Copy' and 'Paste' to a new Microsoft Word document.

4. Edit in MS Word if required to add text, choose aberrations etc.
5. The report includes the following information:

    - Image of the aberration with flanking probes
    - Location of the aberration
    - P-value (as discussed in section 5.8 below)
    - The number of probes in the aberration (#)
    - If the aberration is in a Syndromic region (and OMIM link)
    - If the aberration is in the CNV database (DGV database in Toronto)
    - If the aberration is in a Recombination Hotspot region (as defined by Bailey *et al*)
    - Genes within the aberration (with Ensembl link)
    - The genes located within the aberration
    - 'Important genes'. The program searches the gene against a list of known human disease genes and any found will be highlighted as an' important' gene
    - Space to add notes.

    Customising the report

The content of the report can be changed according to the user requirements. To do this select Tools>Options>Report generation

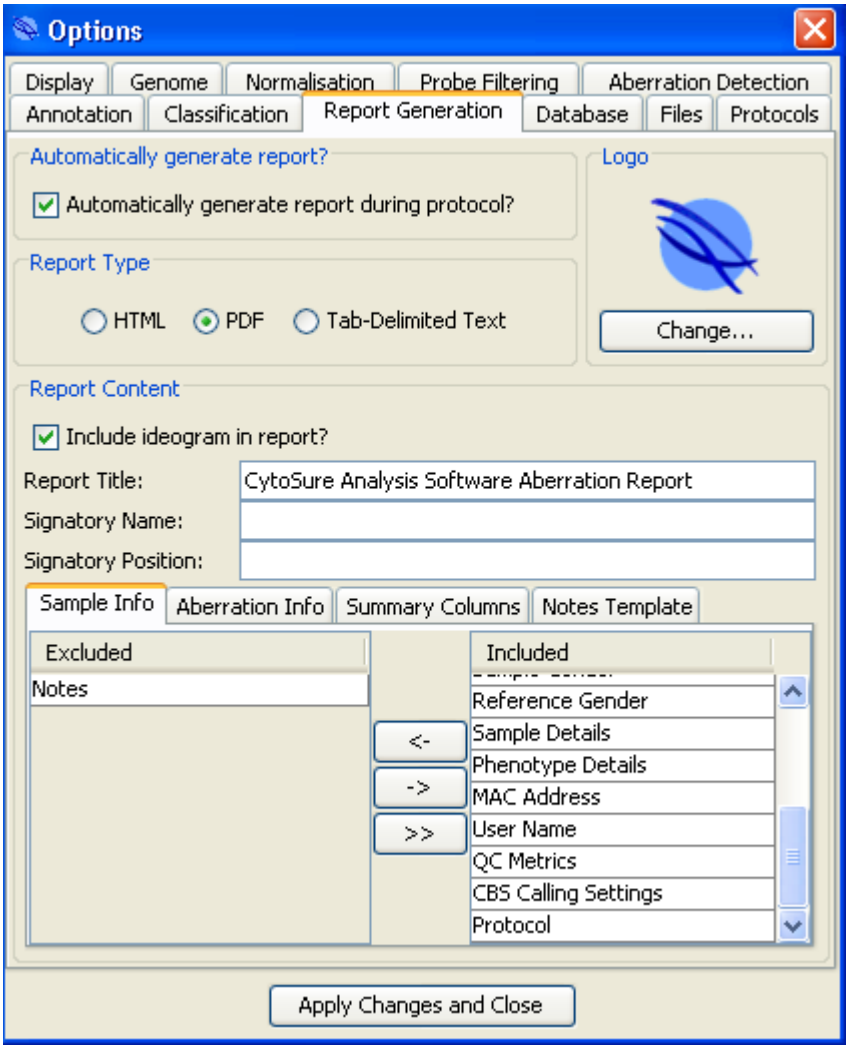


Figure 61 – Customizing the report

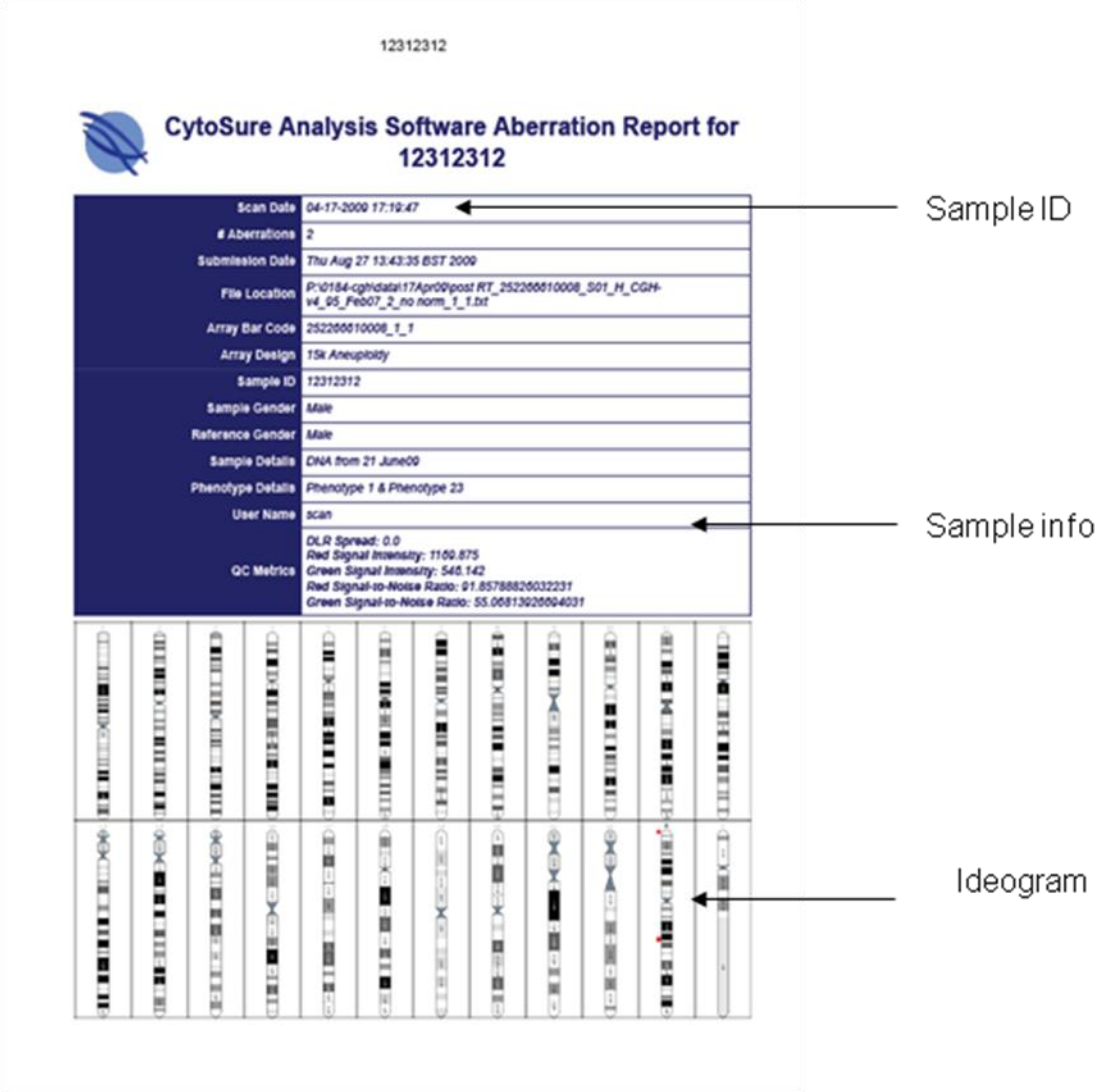
The HTML and pdf reports have a similar format. The HTML files can be opened in Microsoft Word for further editing.
A logo can be added to the report by selecting the 'change' button and by navigating to a file containing the relevant logo. The files must be in JPEG or BMP format.

The information included in the Sample Info, Aberration Info and Summary columns are all user definable

Information can be added or removed from the report by highlighting the relevant information and selecting the -> or <- button.

The Notes Template allows the user to add relevant text which will appear on all the reports.

The layout of the report is shown in the Figures below

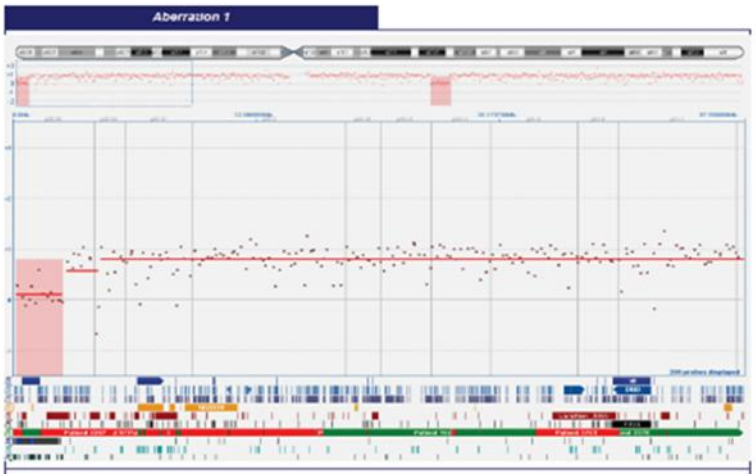

12312312
CytoSure Analysis Software Aberration Report for 12312312
Scan Date 04-17-2009 17:19:47
Aberrations 2
Submission Date Thu Aug 27 13:43:35 BST 2009
File Location P:\0184-cgh\data\17Apr09\post RT_252266510008_S01_H_CGH-v4_65_Feb07_2_no norm_1_1.txt
Array Bar Code 252266510008_1_1
Array Design 15k Aneuploidy
Sample ID 12312312
Sample Gender Male
Reference Gender Male
Sample Details DNA from 21 June09
Phenotype Details Phenotype 1 & Phenotype 23
User Name scan
QC Metrics DLR Spread: 0.0
Red Signal Intensity: 1150.875
Green Signal Intensity: 545.142
Red Signal-to-Noise Ratio: 91.85788826032231
Green Signal-to-Noise Ratio: 55.06813920594031
Sample ID
Sample info
Ideogram

12312312

	Chr	Cytogenetic Location	Start	Stop	Size (Mb)	P-Value	Gain/Loss
1	X	p22.33	164878	2524916	2.36	0.0045	Loss
2	X	q21.31-q21.32	88476955	92191463	3.714	0.00285	Loss

Summary columns

12312312



Chr : X
Start : 164878
Stop : 2524916
Cytogenetic Location : p22.33
Size (Mb) : 2.36
Probes : 15
Classification : Unclassified
CNV % : 61%
Gain/Loss CNV % : 3%
DB CNV % : 0%
DB Gain/Loss CNV % : 0%
CNV Regions : Variation_33160, Variation_23328, Variation_33161, Variation_33162
Syndrome Region? : SHOX, Leri-Weill dyschondrosteosis (LWD) - SHOX deletion
Important Genes : CSF2RA, SHOX
All Genes : GTPBP6, PPP2R3B, RP13-465B17.2, RP13-465B17.3, SHOX, RP11-309M23.2, CRLF2, CSF2RA, 5S_rRNA, IL3RA, SLC25A6, ASMTL, P2RY8, CXYorf3, ASMT, SSU_rRNA_5, DHRSX, BX649443.16, ZBED1
Decipher Regions : Patient 978, Patient 248328, Patient 2207

Aberration info

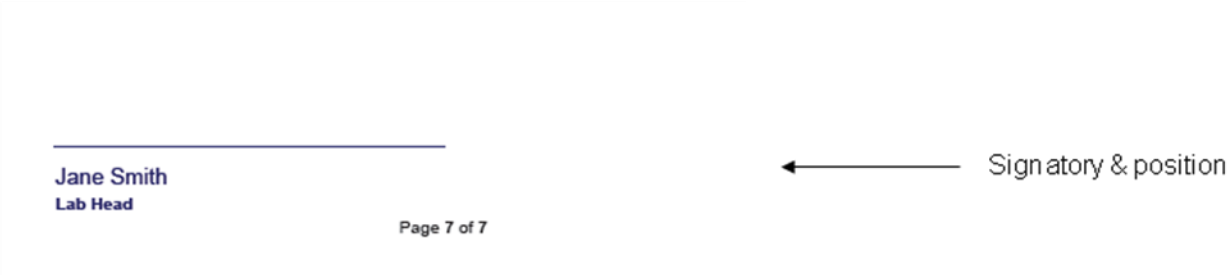


Figure 62 – A custom report

5.7 Exporting the aberration results in Decipher format

When the list of aberrations has been generated and edited it can also be saved in a format that is ready to be uploaded to Decipher simply by selecting the button “Export the Aberration List in Decipher Format” visible in figure 48.

5.8 P-value

1. The method of calculating the P-value involves working out the difference in means between the probes in the aberrations, which have been highlighted by the user, and the mean in the control region. This value is then divided by the standard deviation of the control region. The P-value is then calculated assuming a normal distribution.

2. The control region comprises all probes in the same chromosome that have not been highlighted as an aberration by the user.

3. **Therefore it is important to highlight all the aberrations in the chromosome to ensure that the control region is correct.**

4. There are many different statistical tests that can be carried out. The test used here is a conservative test; it does not take account of the number of probes in the aberration. The P-value can be used as an indication of the quality of the result. However, it is important not to place too much emphasis on the P-value and also consider the number of probes in the aberration. Other tests such as the t test or Mann-Whitney test frequently give much lower P-values on the same datasets.

If there are two datasets (e.g. dye swap) the software uses both datasets to calculate the P-value.

6. Loop experimental design

An alternative experimental design to the reference design is the loop design. This uses 3 arrays for 3 samples and results in duplicate data for each sample. Hybridizations are carried out as follows:

Array 1: Sample 1 versus Sample 2
Array 2: Sample 2 versus Sample 3
Array 3: Sample 3 versus Sample 1

CytoSure software supports this experimental design.

There are two different methods available to analyse the loop design.

The first method uses the CBS calls of all 3 datasets. In this method each array is analysed and the aberrations are called. The aberrations are then combined

The workflow is as follows:

Step 1: Import the 3 arrays and run the CBS
Step 2: Check the aberration called for each sample
Step 3: Run the loop analysis
Step 4: Manually examine the replication
Step 5: Rerun the loop analysis

This method is described in figure 6.1, 6.2, 6.3 and 6,4

The second method combines the 2 relevant replicate datasets before calling the aberrations.

The workflow is as follows:

Step 1: Import the 3 arrays
Step 2: Run the loop analysis and combine the replicates
Step 3: Examine the called aberrations

This combination method is described in Section 6.5.

6.1 Import the 3 arrays and run the CBS

The 3 feature extracted files (.txt from Agilent scanner, gpr from Axon scanner) are imported into CytoSure software using File -> Import (see section 3). The Cy5 sample details are inputted. The Cy3 samples details do not need to be inputted at this stage (figure 49)

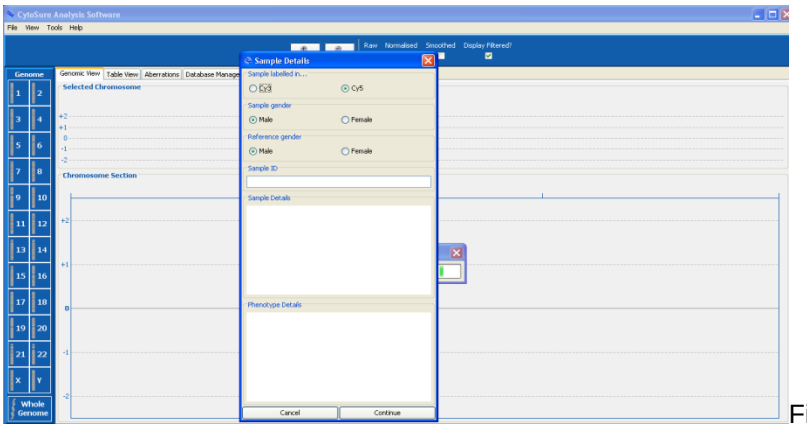


Figure 62 - Import of the array data.

Once the 3 datasets are imported into CytoSure software, the CBS can be run for each dataset. Please see section 5.1 for more details.

6.2 Check the aberrations manually

The process of manually checking the aberrations is easier to do if 1 dataset is checked at a time. Alterations to the list of aberrations can be done using the manual aberration tools (see section 5.2). It is necessary to use the tabs on the aberration table to toggle between the datasets (see figure 63).

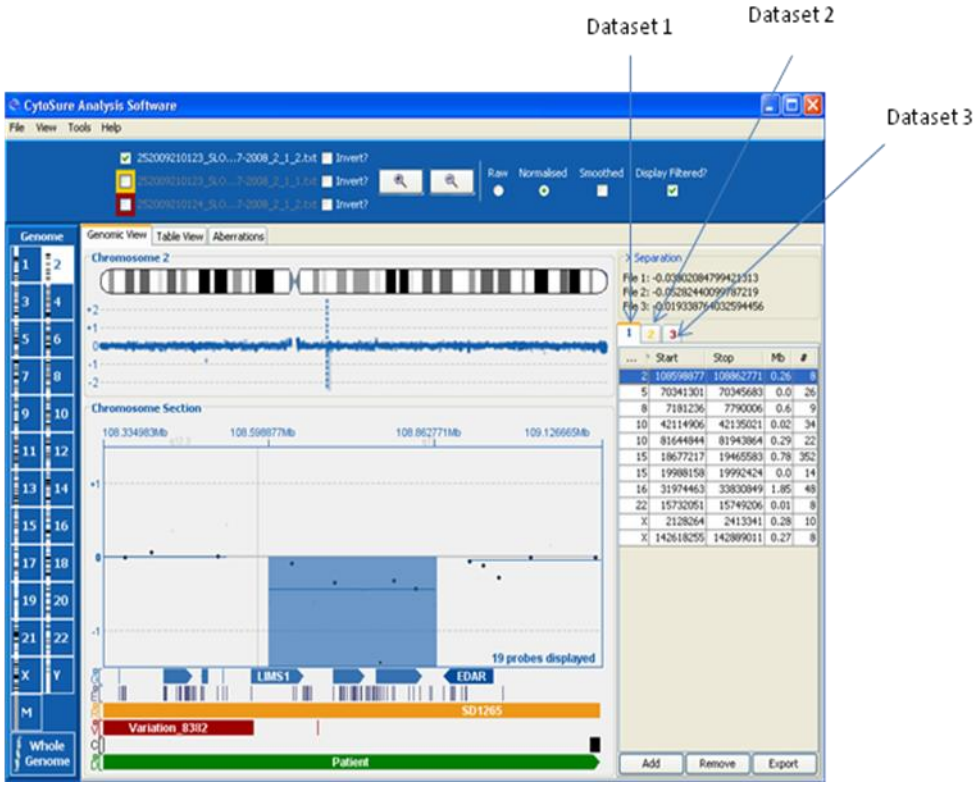


Figure 63 - Checking the automated calling.

6.3 Running the loop analysis

Click Tools -> Run analysis

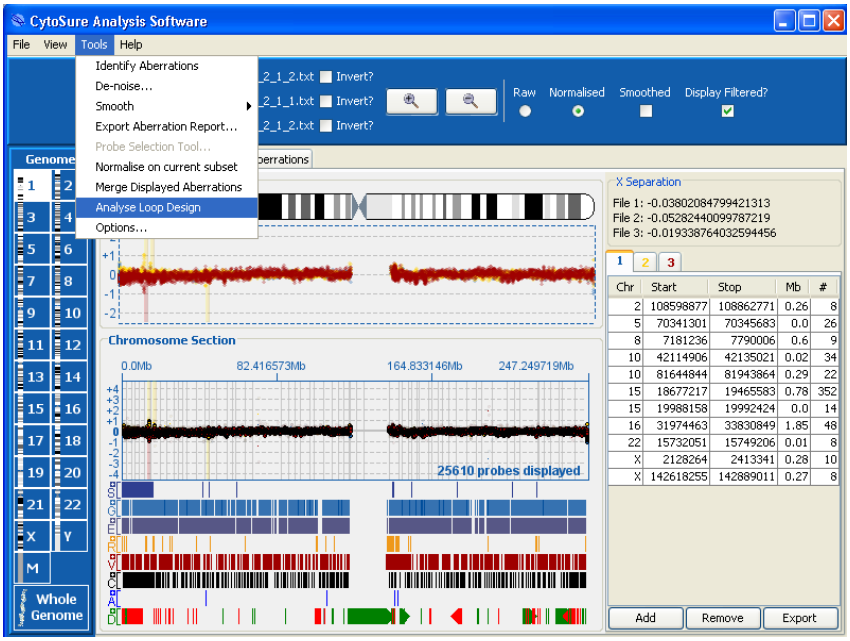


Figure 64 - Running the loop analysis.

A dialog box will appear that will allow the user to input the sample details and with which dye the samples were labelled. If some of the sample details were inputted during file input then some of these details will be automatically generated. Under analysis method click ‘Use CBS calls’ (see figure 65).

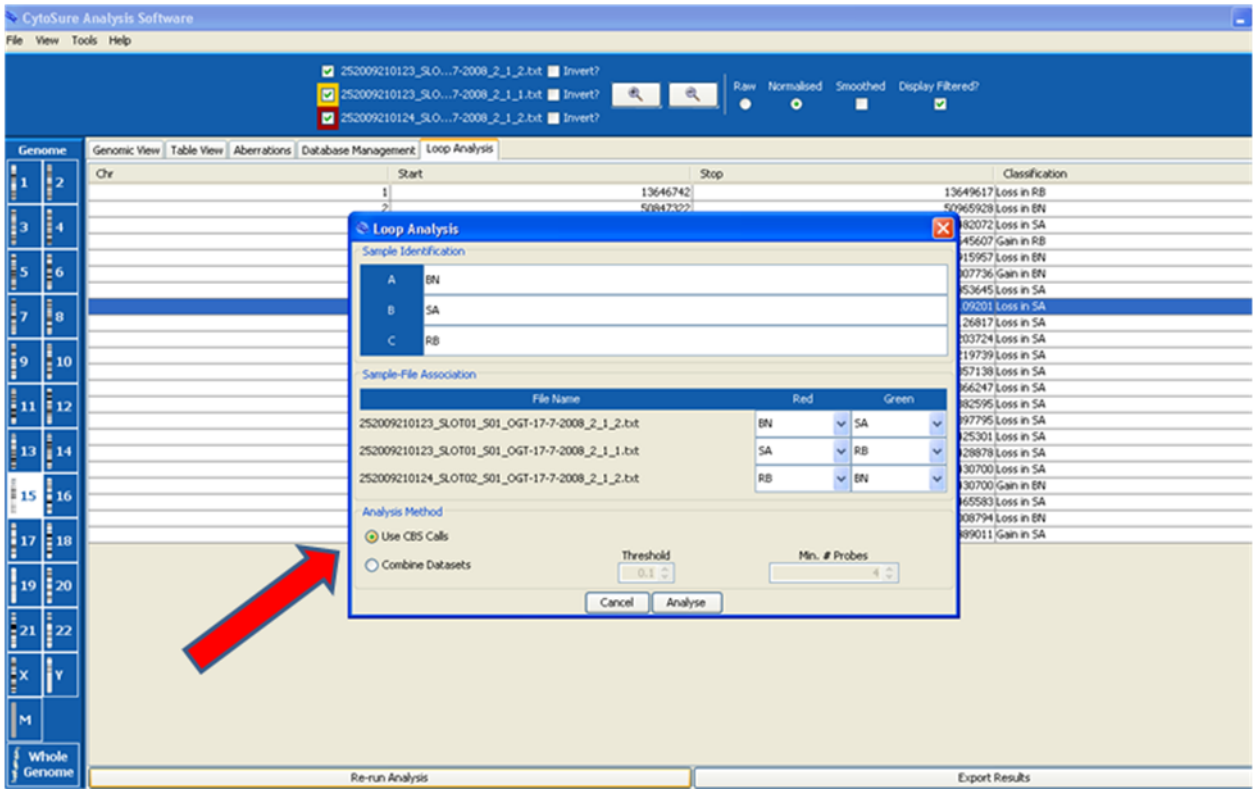


Figure 65 - Sample input in loop analysis and selection of the use CBS call analysis method.

An additional tab will appear. Clicking on this tab will display the loop analysis page (figure 66). This displays the position of the aberrations in all 3 datasets and will classify the aberrations according to whether they are present in the replicate array results. If the aberrations have replicated in the two datasets exactly then the aberration will be classified definitively (e.g. loss in A, gain in B). However, if there are differences between the 2 datasets then aberrations will be classified ambiguously (e.g. Gain in A or Loss in B).

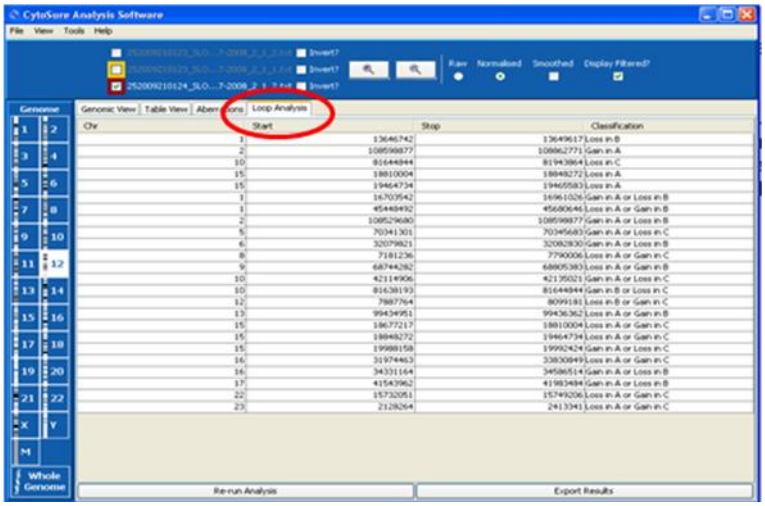


Figure 66 - The loop analysis page.

6.4 Manually examine replicates

Click on each aberration and the aberration will be displayed in genomic view. Figure 67 shows an aberration which has been replicated exactly.

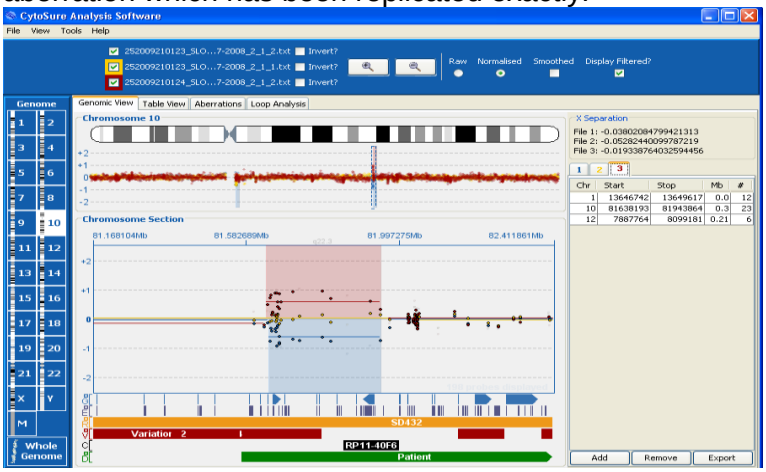


Figure 54 - An aberration that has replicated exactly.

There are 3 possible reasons for the aberrations not to be called identically in the replicates.

The first reason is that the automated aberration detection may not have called the aberration in one dataset. This is shown in figure 68 where an aberration has been called in dataset 1 (the blue dataset) but not in dataset 2 (the yellow dataset)

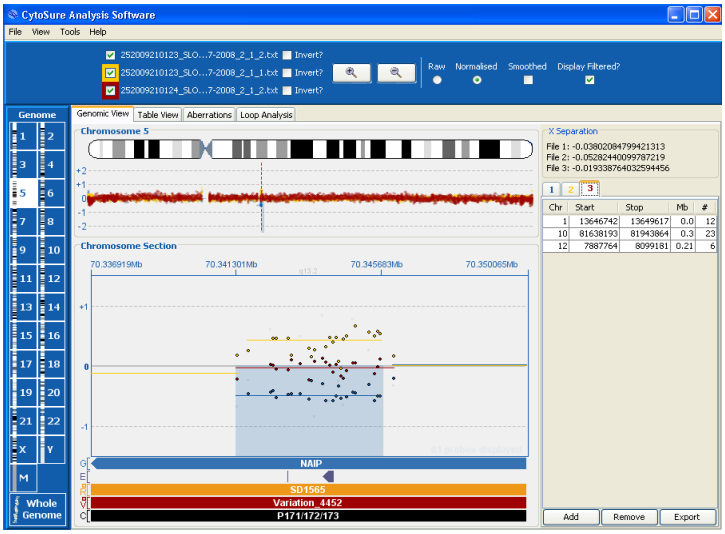


Figure 68 - One dataset is not called.

If the user wants to call the yellow dataset as the blue dataset then in this example right click on the blue block and select ‘add matching aberration’.

A second reason for differences in the replication is due to different sizes of aberration called (figure 69).

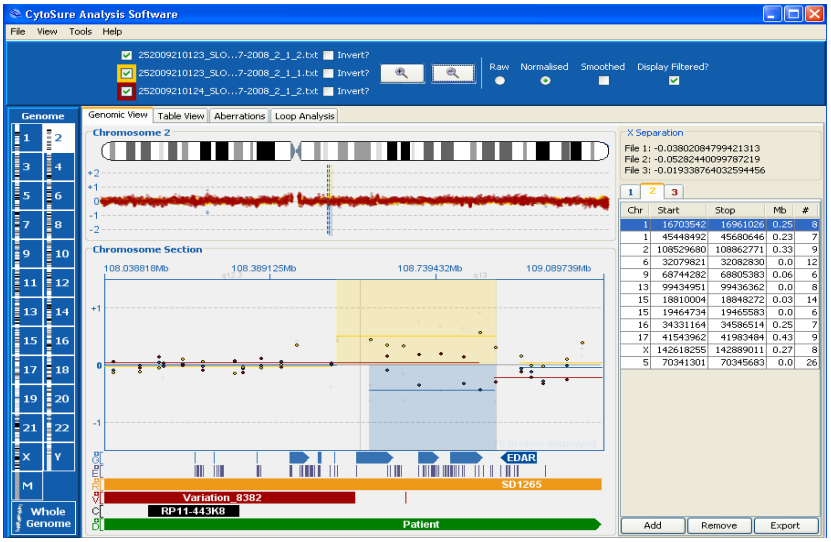


Figure 69- The aberrations are called with different sizes.

In this example, if need to alter the dataset 1 (blue) aberration to the size of the dataset 2 (yellow) aberration, then the first step is to delete the blue aberration by right clicking on the blue block and selecting delete. Then right click on the yellow block and select ‘add matching dataset.

The final reason might be that the automated aberration detection has called a questionable aberration (see figure 70). If this is the case then right click on the block and select delete.

Figure 70 - A questionable aberration.

6.5 Re-run the loop analysis

Once the changes have been made to the aberration lists, then rerun the loop analysis, as in section 6.3. This will refresh the loop analysis table.

6.6 Saving aberrations to the database

To save the aberrations in the database, click on the aberration tab to view all the aberrations The aberration view will list the aberrations from the 3 arrays, both the original CBS calls and also the loop analysis results. Therefore, if the user wants to save in the database only the aberrations from the loop analysis, click on the drop down menu (see figure below). Select Automatic-CBS and then ‘Remove all aberrations of the selected type’. The aberrations can then be stored in the database as described in section 5.4.

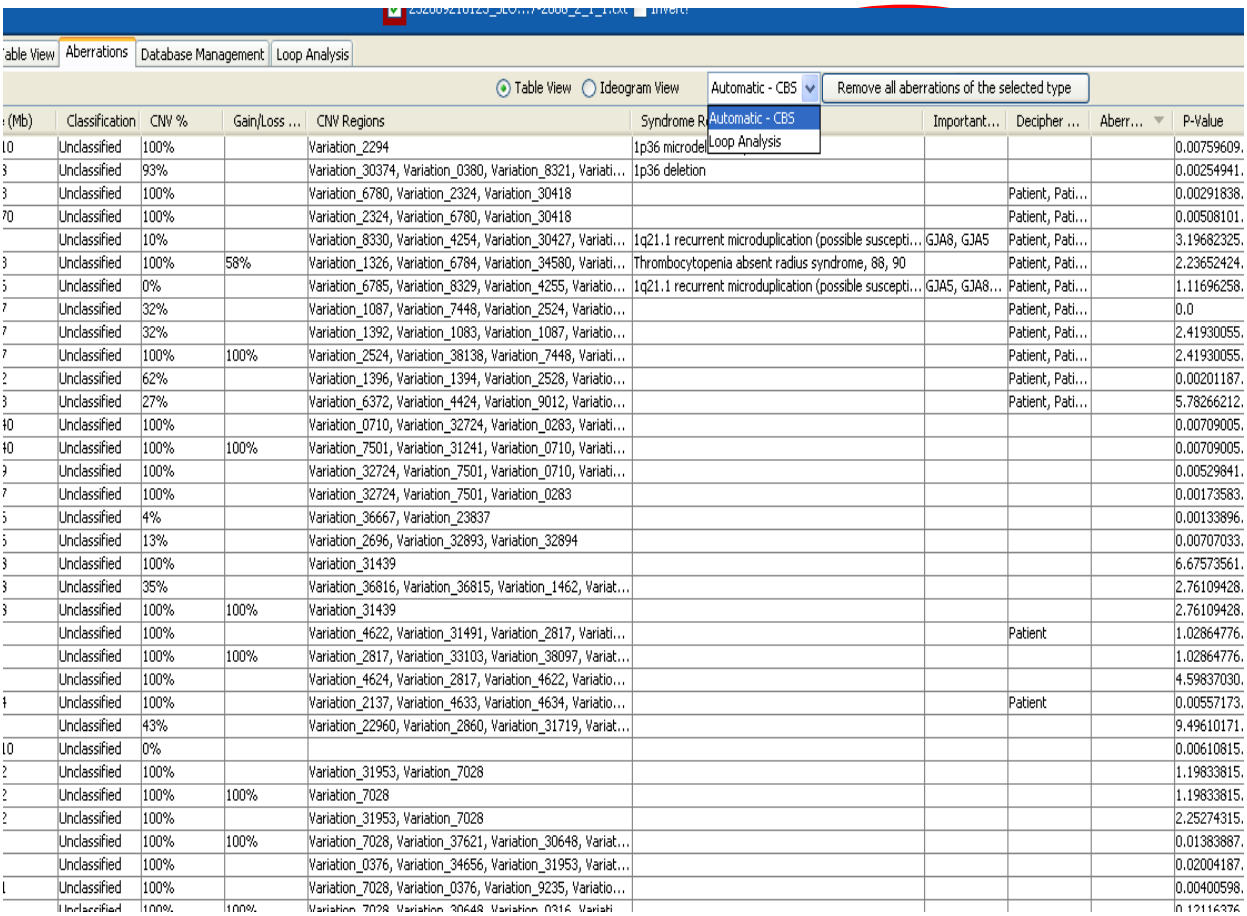

able View
Aberrations
Database Management
Loop Analysis
Table View
Ideogram View
Automatic - CBS
Remove all aberrations of the selected type
Automatic - CBS
Loop Analysis
(Mb)
Classification
CNV %
Gain/Loss ...
CNV Regions
Syndrome R
Important...
Decipher ...
Aberr...
P-Value
Unclassified 100% Variation_2294 1p36 microdel 0.00759609.
Unclassified 93% Variation_30374, Variation_0380, Variation_8321, Variati... 1p36 deletion 0.00254941.
Unclassified 100% Variation_6780, Variation_2324, Variation_30418 Patient, Pati... 0.00291838.
Unclassified 100% Variation_2324, Variation_6780, Variation_30418 Patient, Pati... 0.00508101.
Unclassified 10% Variation_8330, Variation_4254, Variation_30427, Variati... 1q21.1 recurrent microduplication (possible suscepti... GJA8, GJA5 Patient, Pati... 3.19682325.
Unclassified 100% 58% Variation_1326, Variation_6784, Variation_34580, Variati... Thrombocytopenia absent radius syndrome, 88, 90 Patient, Pati... 2.23652424.
Unclassified 0% Variation_6785, Variation_8329, Variation_4255, Variatio... 1q21.1 recurrent microduplication (possible suscepti... GJA5, GJA8... Patient, Pati... 1.11696258.
Unclassified 32% Variation_1087, Variation_7448, Variation_2524, Variatio... Patient, Pati... 0.0
Unclassified 32% Variation_1392, Variation_1083, Variation_1087, Variatio... Patient, Pati... 2.41930055.
Unclassified 100% 100% Variation_2524, Variation_38138, Variation_7448, Variati... Patient, Pati... 2.41930055.
Unclassified 62% Variation_1396, Variation_1394, Variation_2528, Variatio... Patient, Pati... 0.00201187.
Unclassified 27% Variation_6372, Variation_4424, Variation_9012, Variatio... Patient, Pati... 5.78266212.
Unclassified 100% Variation_0710, Variation_32724, Variation_0283, Variati... 0.00709005.
Unclassified 100% 100% Variation_7501, Variation_31241, Variation_0710, Variati... 0.00709005.
Unclassified 100% Variation_32724, Variation_7501, Variation_0710, Variati... 0.00529841.
Unclassified 100% Variation_32724, Variation_7501, Variation_0283 0.00173583.
Unclassified 4% Variation_36667, Variation_23837 0.00133896.
Unclassified 13% Variation_2696, Variation_32893, Variation_32894 0.00707033.
Unclassified 100% Variation_31439 6.67573561.
Unclassified 35% Variation_36816, Variation_36815, Variation_1462, Variat... 2.76109428.
Unclassified 100% 100% Variation_31439 2.76109428.
Unclassified 100% Variation_4622, Variation_31491, Variation_2817, Variati... Patient 1.02864776.
Unclassified 100% 100% Variation_2817, Variation_33103, Variation_38097, Variat... 1.02864776.
Unclassified 100% Variation_4624, Variation_2817, Variation_4622, Variatio... 4.59837030.
Unclassified 100% Variation_2137, Variation_4633, Variation_4634, Variatio... Patient 0.00557173.
Unclassified 43% Variation_22960, Variation_2860, Variation_31719, Variat... 9.49610171.
Unclassified 0% 0.00610815.
Unclassified 100% Variation_31953, Variation_7028 1.19833815.
Unclassified 100% 100% Variation_7028 1.19833815.
Unclassified 100% Variation_31953, Variation_7028 2.25274315.
Unclassified 100% 100% Variation_7028, Variation_37621, Variation_30648, Variat... 0.01383887.
Unclassified 100% Variation_0376, Variation_34656, Variation_31953, Variat... 0.02004187.
Unclassified 100% Variation_7028, Variation_0376, Variation_9235, Variatio... 0.00400598.

Figure 71 - Removing aberrations so only the loop data remains.

6.7 Alternative method: the combination method

An alternative method to carrying out a loop analysis is to combine the dataset prior to doing the analysis.

To do this click Tools ->Loop analysis. The loop analysis box opens. A dialog box will appear that will allow the user to input the sample details and to specify with which dye the samples were labelled. If some of the sample details were inputted during file input then some of these details will be automatically generated.

Under analysis method click ‘Combine datasets’ (step 1 in figure 72), then set the Threshold (step 2 in figure 72) and the minimum number of probes (step 3 in figure 72).

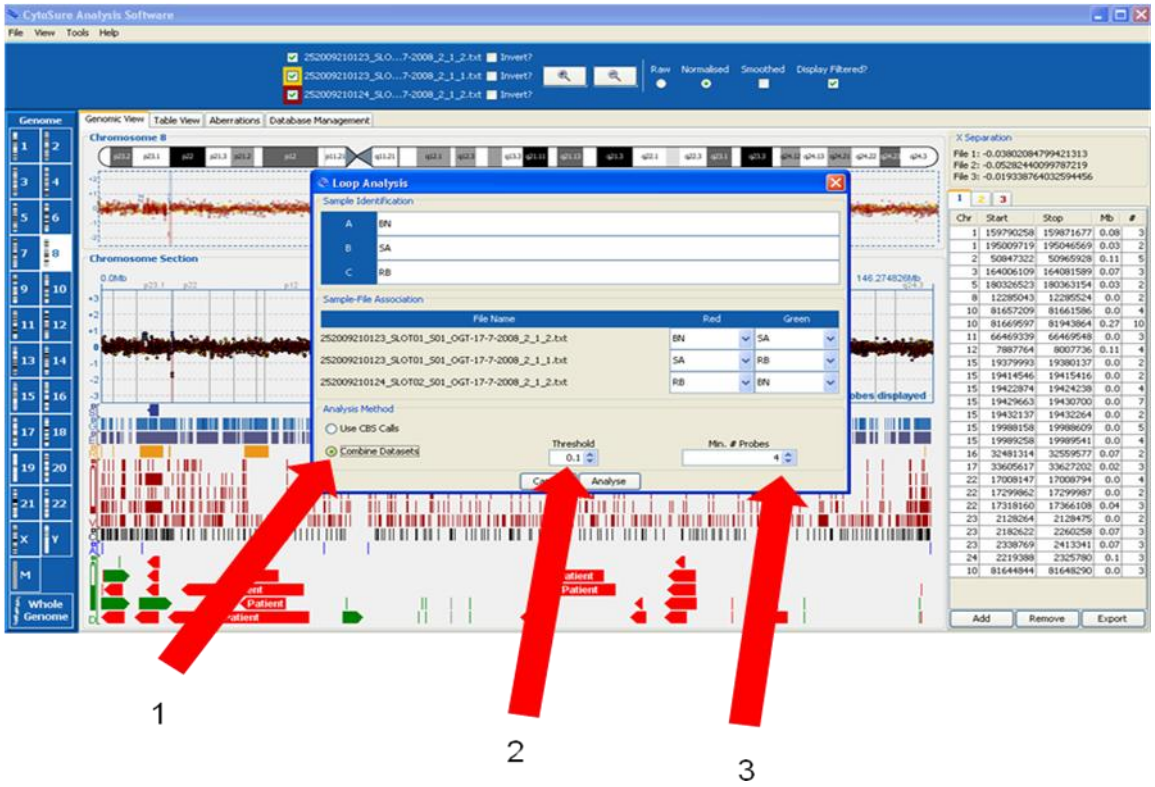


Figure 72 - Using the method of combining the datasets before analysis.

The software will combine the datasets in such a way that the aberrations that have been called correctly by both arrays will have a high value. Those that have only been called in one dataset will have a low value. The results are then outputted as shown below:

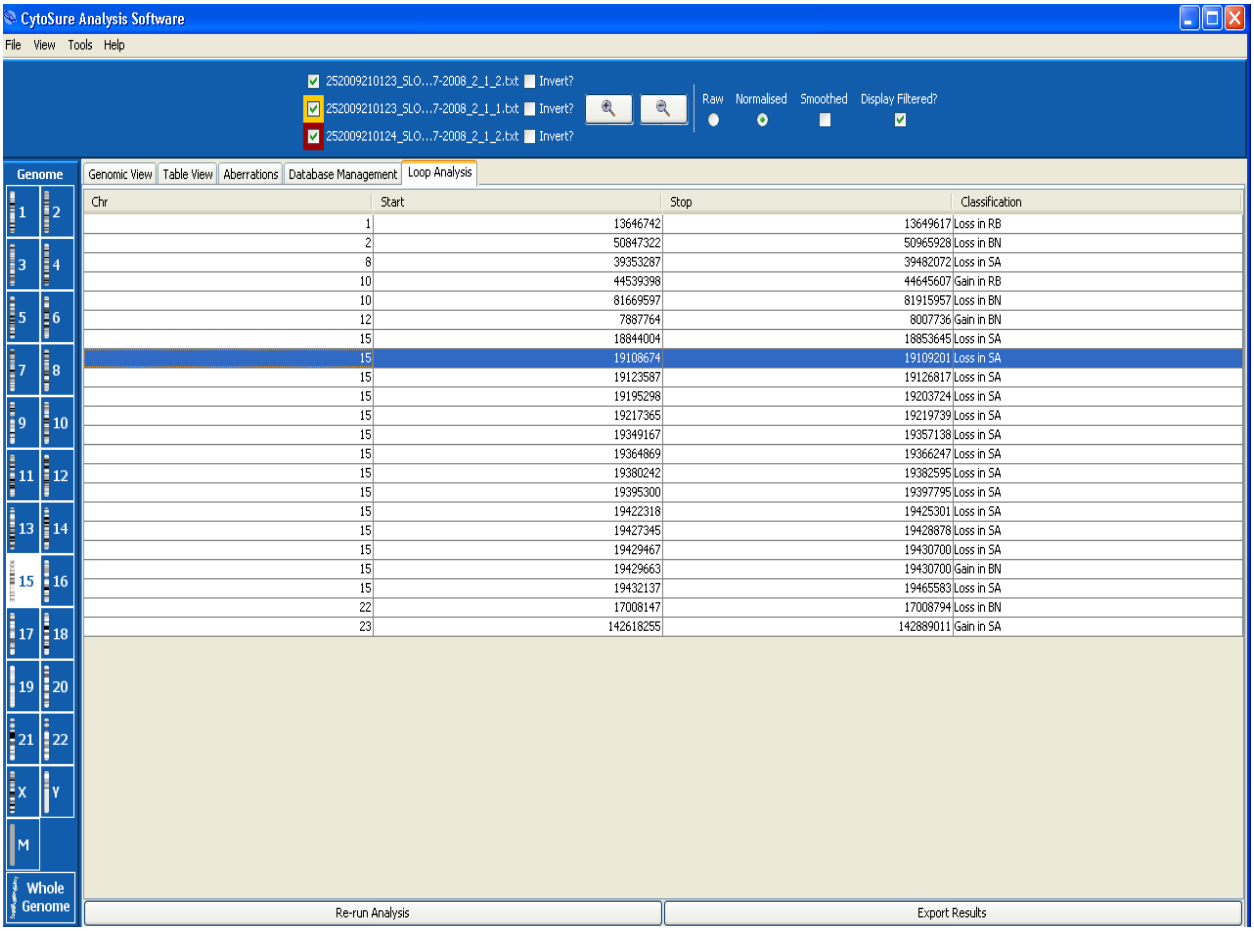


Figure 73: Output from the loop analysis using the combination method.

6.8 Saving aberrations to the database when using combination tool

Click on aberration tab. If the data had previously been analysed by CBS, then these aberrations can be rapidly removed.

7. Database management

CytoSure analysis software enables aberrations called and annotations added (see process described in Chapter 5) to be saved to a database. Aberrations saved are recalled on a custom track in the genome view (the blue track labelled 'A', see section 4.2). This section describes how to manipulate the database, for example, to edit or delete data after saving.

7.1 Positioning the database file in a location of user's choice

The data is stored as a db.xml file which by default is located at C:\ Program Files\ CytoSure Analysis Software. This file can be moved using Windows explorer to a different drive e.g. a shared drive. *The file should be on a drive that is backed up regularly.* If this is done then the database needs to be linked to the db.xml file as follows:

1. Go to the database modification page by clicking on the database modification tab.

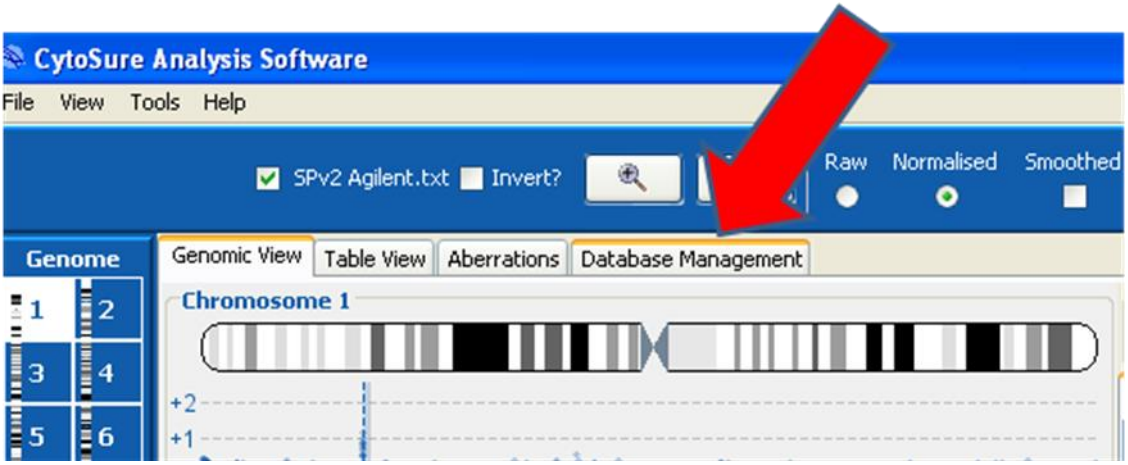


Figure 76 - Opening database modification page.

2. Click on the File address at the top left hand side of the screen (figure 77).

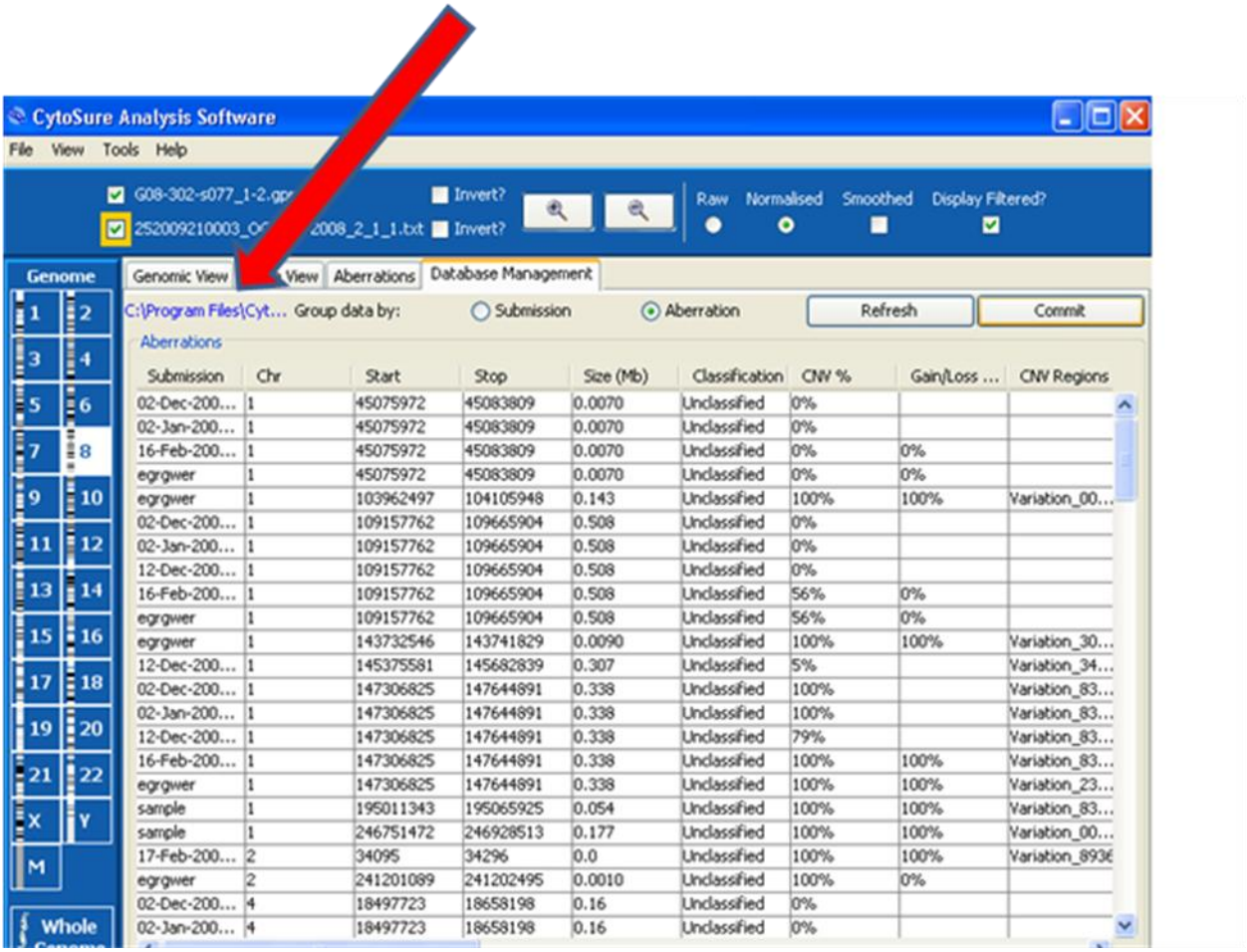


Figure 76 - Mapping the db.xml file location.

3. Navigate to position of the moved db.xml file and click open

7.2 Viewing data in the database

Click on the database management tab to access information stored in the database.

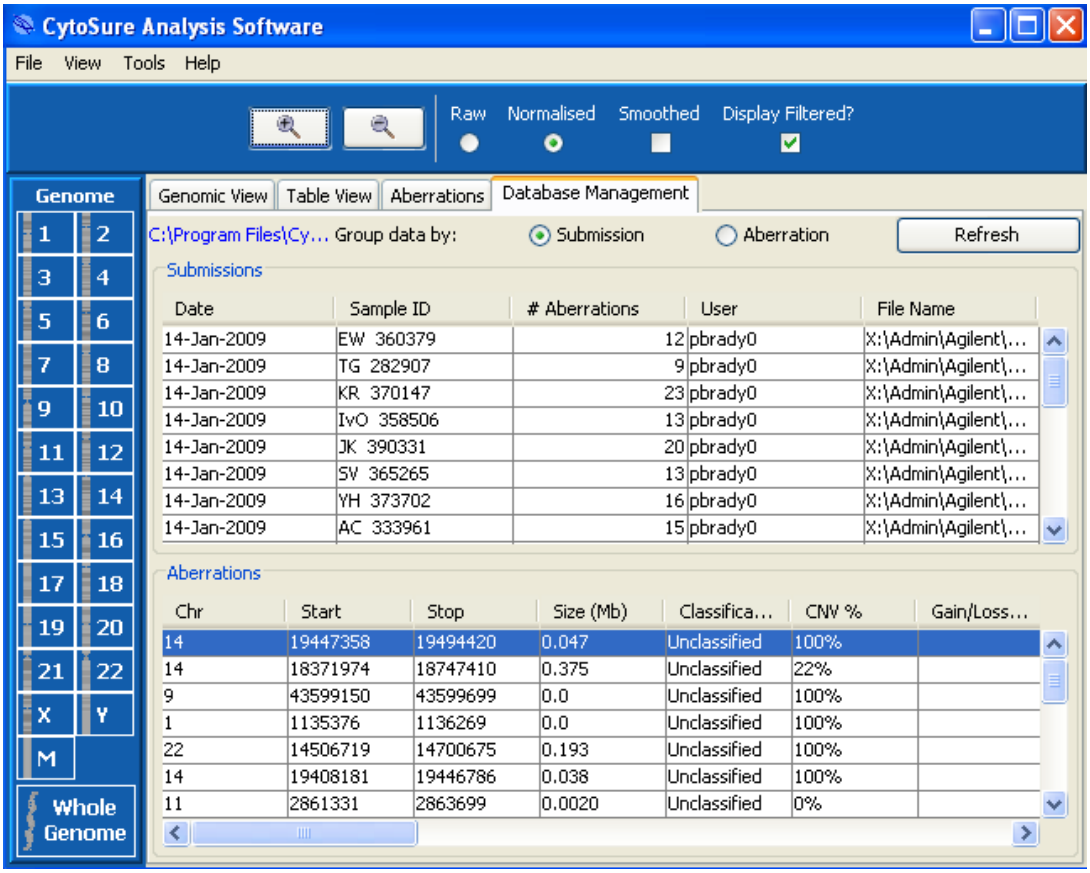

CytoSure Analysis Software
File View Tools Help
Raw Normalised Smoothed Display Filtered?
Genome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y M
Whole Genome
Genomic View Table View Aberrations Database Management
C:\Program Files\Cy... Group data by: Submission Aberration Refresh
Submissions
Date Sample ID # Aberrations User File Name
14-Jan-2009 EW 360379 12 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 TG 282907 9 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 KR 370147 23 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 IvO 358506 13 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 JK 390331 20 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 SV 365265 13 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 YH 373702 16 pbrady0 X:\Admin\Agilent\...
14-Jan-2009 AC 333961 15 pbrady0 X:\Admin\Agilent\...
Aberrations
Chr Start Stop Size (Mb) Classifica... CNV % Gain/Loss...
14 19447358 19494420 0.047 Unclassified 100%
14 18371974 18747410 0.375 Unclassified 22%
9 43599150 43599699 0.0 Unclassified 100%
1 1135376 1136269 0.0 Unclassified 100%
22 14506719 14700675 0.193 Unclassified 100%
14 19408181 19446786 0.038 Unclassified 100%
11 2861331 2863699 0.0020 Unclassified 0%

Figure 77: The database management page.

The data that has been saved can be viewed by submission (date when the data was saved to the database) or by aberration (see figure below). If viewed by aberration the data can be sorted, for example by chromosome by clicking on the appropriate column header.

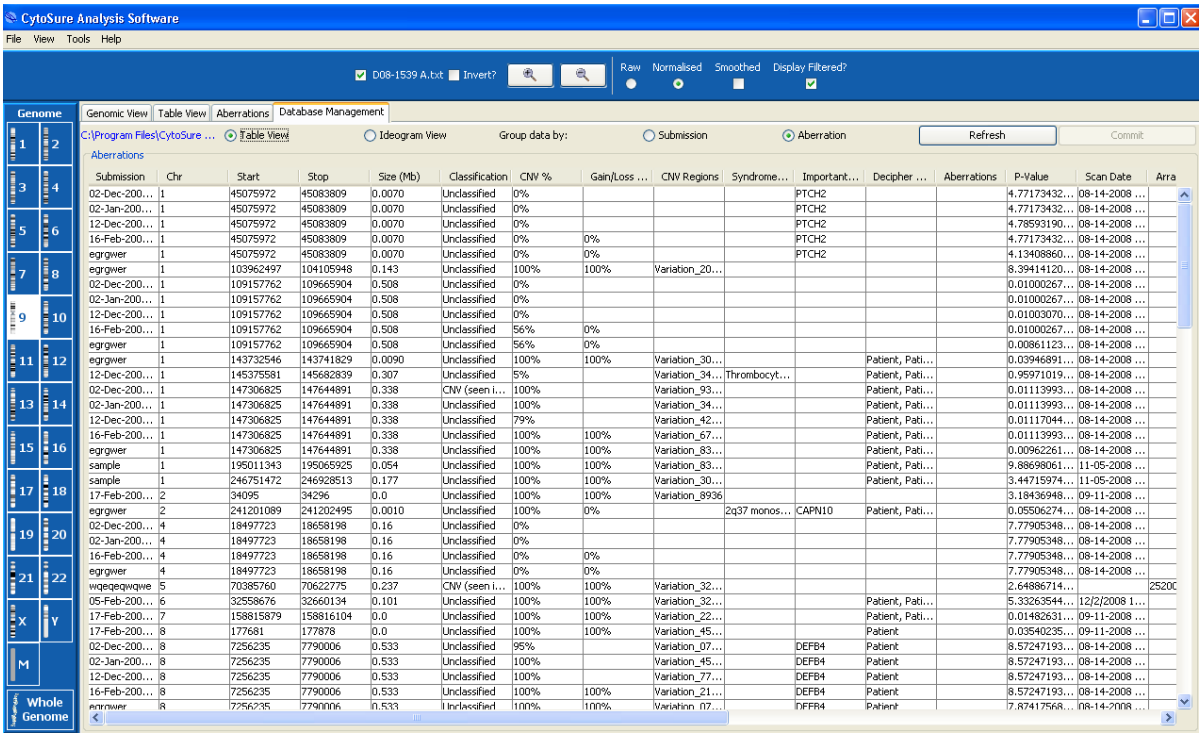


Figure 78 - The data displayed as aberrations. In this case the data has been sorted by clicking on the Chr column heading.

To view the relevant aberration, click on the aberration entry and an image of the aberration in genomic view will be shown (see figure below).

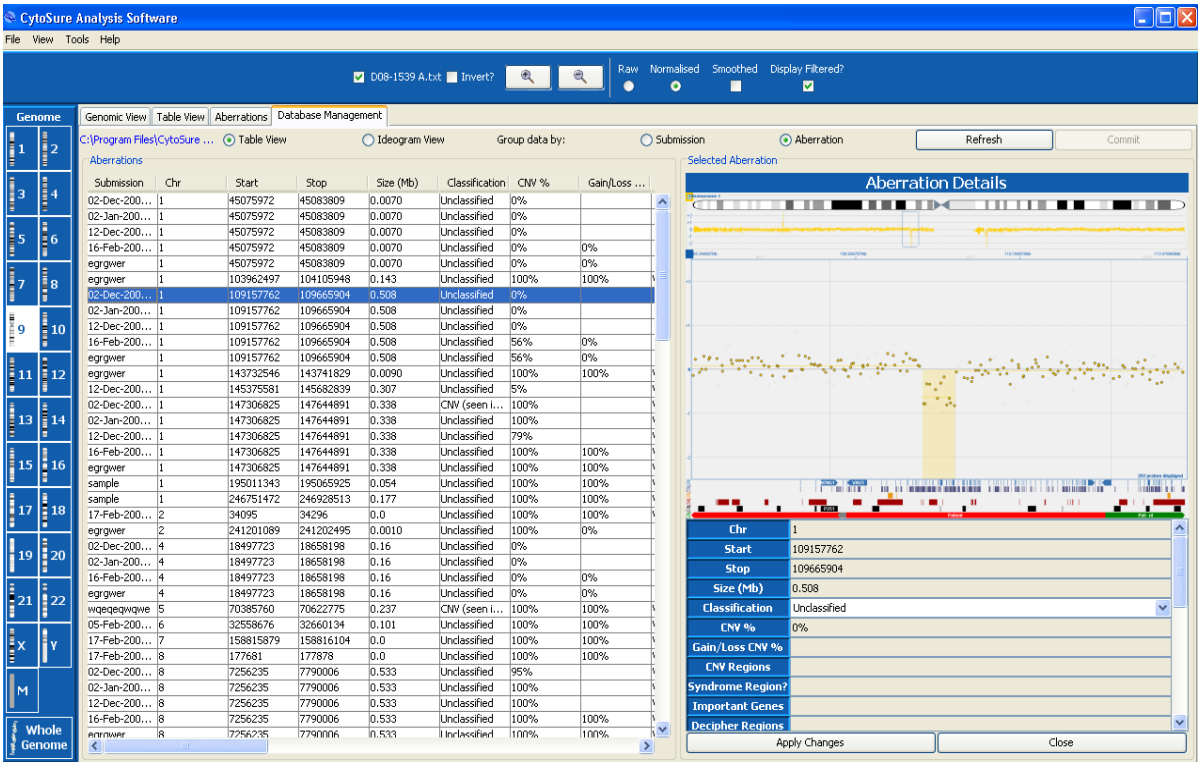


Figure 79 - Detailed image of an aberration stored in the database.

Viewing the data within the database as an ideogram

To display the data in the database as an ideogram click on the relevant radio button as shown in figure 80. This will display the data as an ideogram (figure 81). In this view the green bars represent gains whilst the red bars represent deletions.

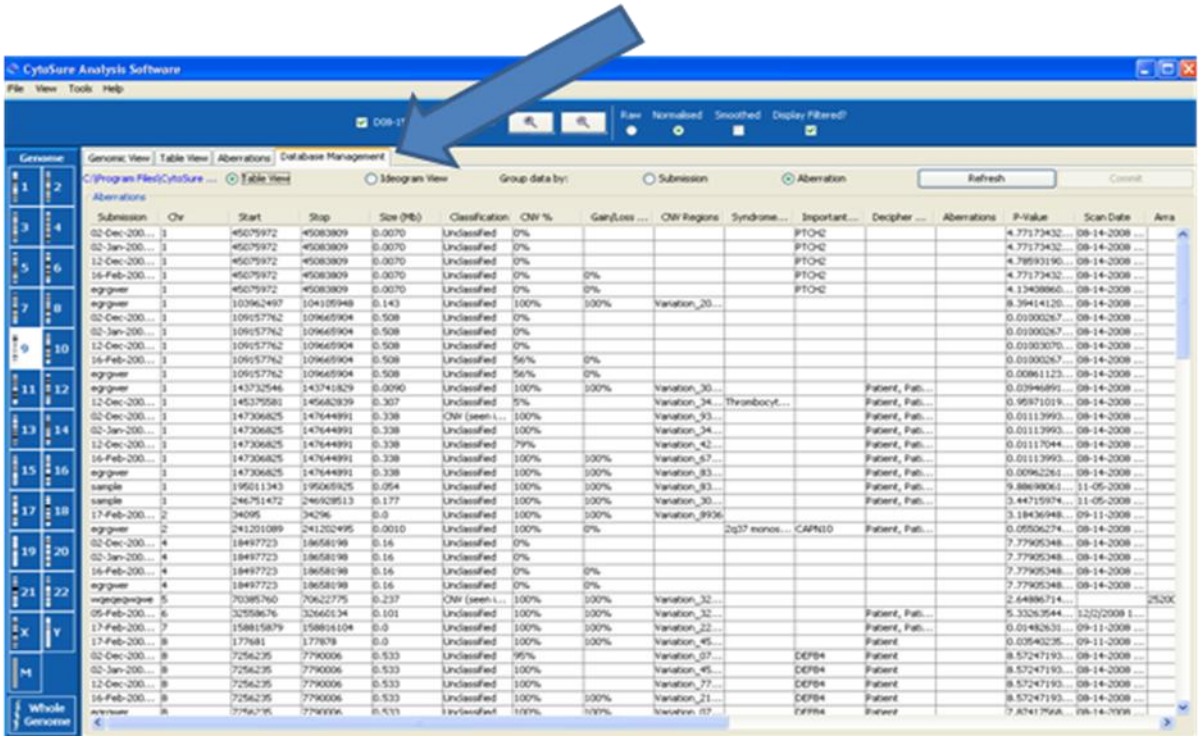


Figure 80 - Radio button to click to display the database information as an ideogram.

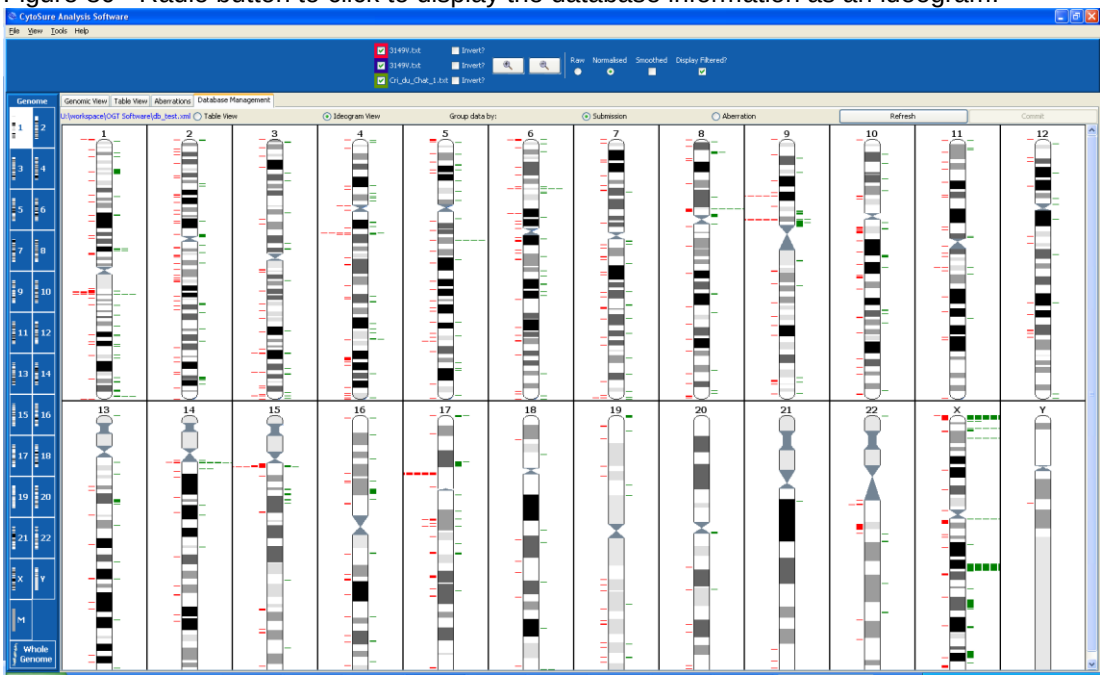


Figure 81 - The aberrations displayed as an ideogram.

For further detail click on a relevant chromosome and the individual chromosome will be displayed, with all the aberrations stored within the database displayed (see figure 82). The display can be returned to the whole genome view by clicking on the chromosome.

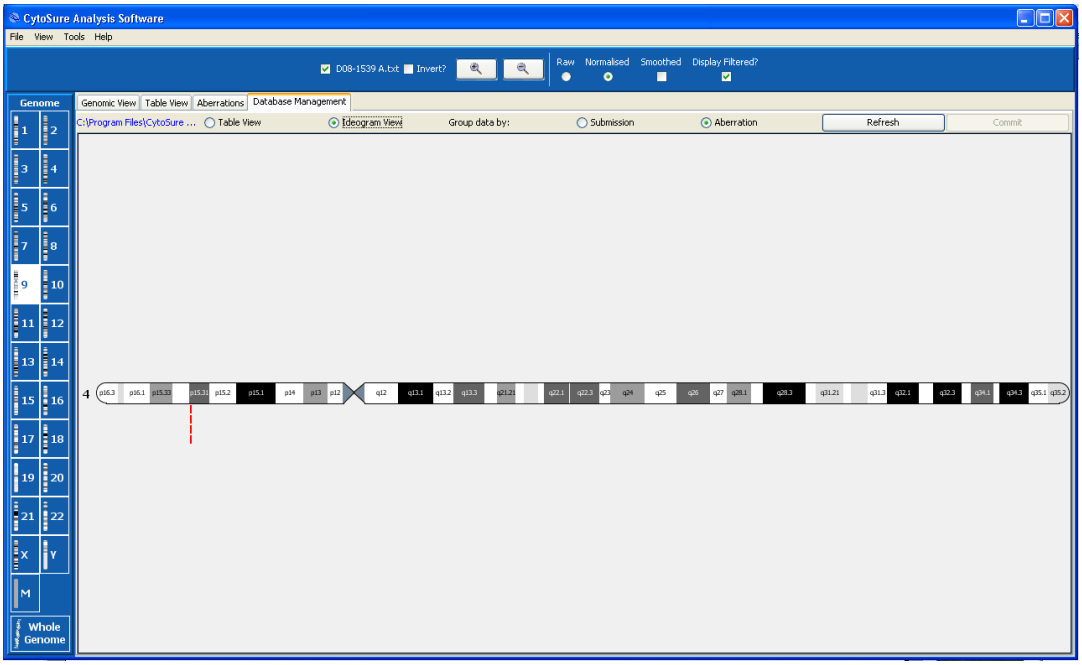


Figure 82 - Individual chromosome view.

7.3 Deleting aberrations in the internal database

To delete a relevant aberration, highlight the relevant data entry and click on the delete key which will bring up a warning (see figure below). In order to delete the aberration, the commit button at the top right of screen needs to be clicked.

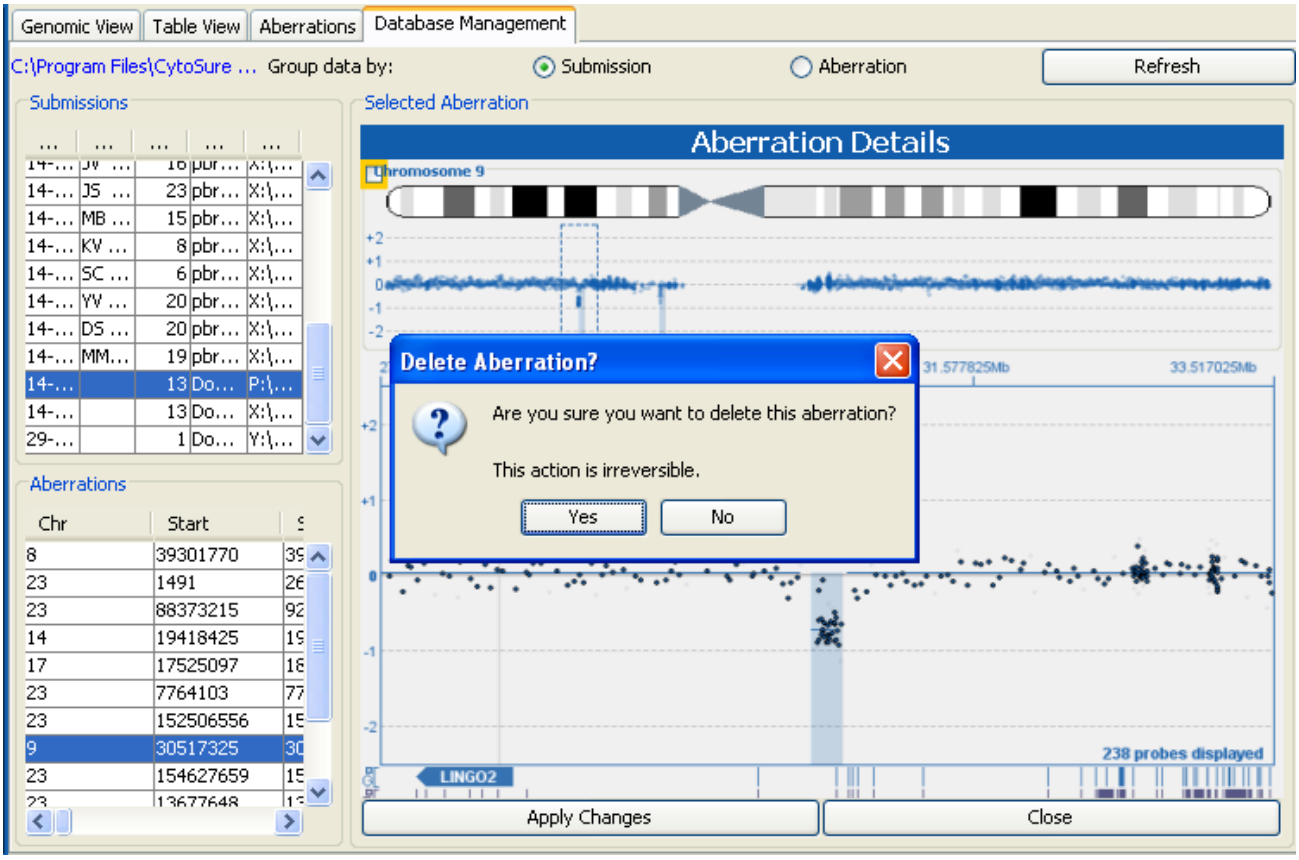


Figure 83 - Deleting an aberration in the database.

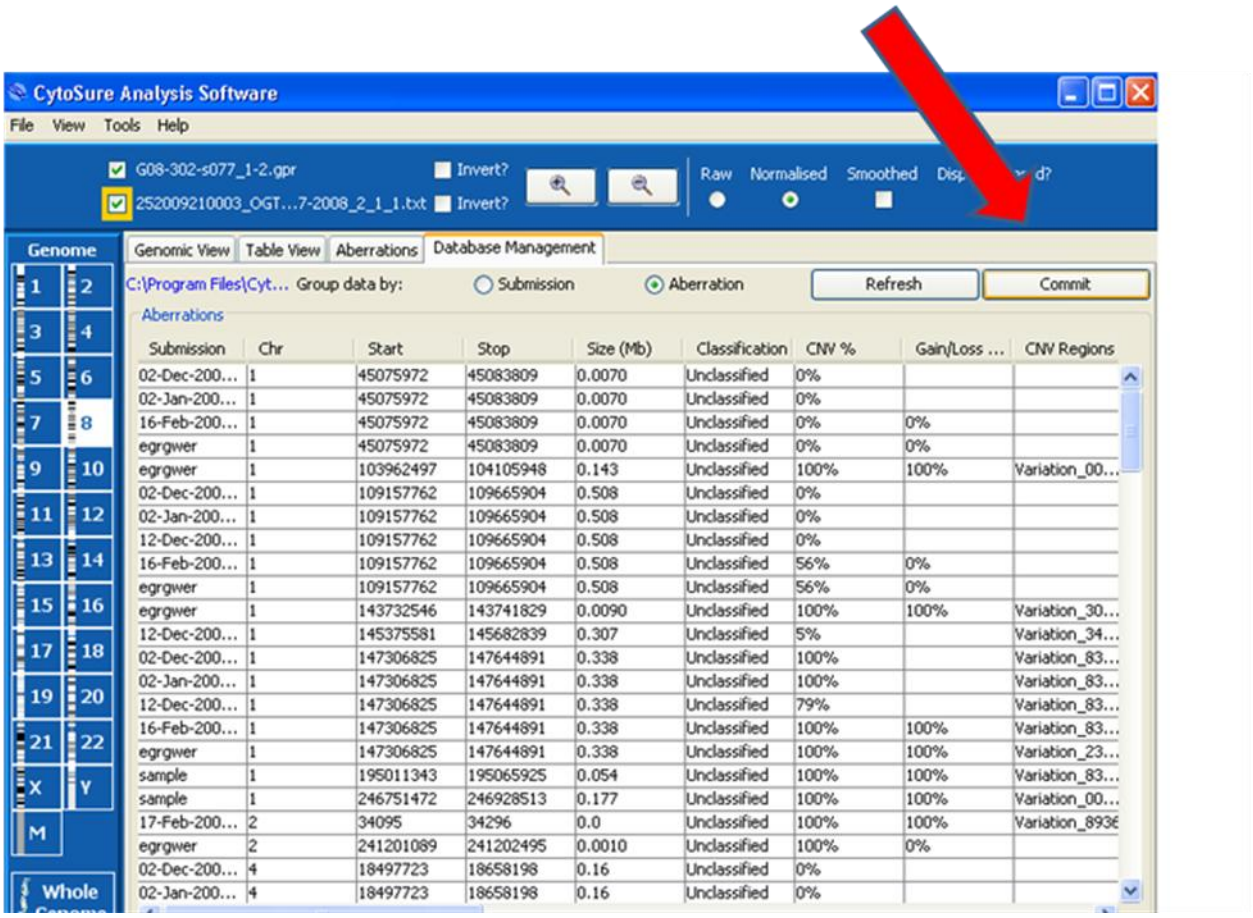

CytoSure Analysis Software
File View Tools Help
G08-302-s077_1-2.gpr
252009210003_OGT...7-2008_2_1_1.txt
Invert?
Raw Normalised Smoothed
Genome
Genomic View Table View Aberrations Database Management
C:\Program Files\Cyt... Group data by:
Submission Aberration
Refresh Commit
Aberrations
Submission Chr Start Stop Size (Mb) Classification CNV % Gain/Loss ... CNV Regions
02-Dec-200... 1 45075972 45083809 0.0070 Unclassified 0%
02-Jan-200... 1 45075972 45083809 0.0070 Unclassified 0%
16-Feb-200... 1 45075972 45083809 0.0070 Unclassified 0% 0%
egrgwer 1 45075972 45083809 0.0070 Unclassified 0% 0%
egrgwer 1 103962497 104105948 0.143 Unclassified 100% 100% Variation_00...
02-Dec-200... 1 109157762 109665904 0.508 Unclassified 0%
02-Jan-200... 1 109157762 109665904 0.508 Unclassified 0%
12-Dec-200... 1 109157762 109665904 0.508 Unclassified 0%
16-Feb-200... 1 109157762 109665904 0.508 Unclassified 56% 0%
egrgwer 1 109157762 109665904 0.508 Unclassified 56% 0%
egrgwer 1 143732546 143741829 0.0090 Unclassified 100% 100% Variation_30...
12-Dec-200... 1 145375581 145682839 0.307 Unclassified 5% Variation_34...
02-Dec-200... 1 147306825 147644891 0.338 Unclassified 100% Variation_83...
02-Jan-200... 1 147306825 147644891 0.338 Unclassified 100% Variation_83...
12-Dec-200... 1 147306825 147644891 0.338 Unclassified 79% Variation_83...
16-Feb-200... 1 147306825 147644891 0.338 Unclassified 100% 100% Variation_83...
egrgwer 1 147306825 147644891 0.338 Unclassified 100% 100% Variation_23...
sample 1 195011343 195065925 0.054 Unclassified 100% 100% Variation_83...
sample 1 246751472 246928513 0.177 Unclassified 100% 100% Variation_00...
17-Feb-200... 2 34095 34296 0.0 Unclassified 100% 100% Variation_8936
egrgwer 2 241201089 241202495 0.0010 Unclassified 100% 0%
02-Dec-200... 4 18497723 18658198 0.16 Unclassified 0%
02-Jan-200... 4 18497723 18658198 0.16 Unclassified 0%
Whole

Figure 84 - Committing changes to the database.

7.4 Editing aberrations in the internal database

In order to edit the annotation associated with an aberration within the database, the following process is used.

1. Navigate to the aberration as described in section 7.1

2. Click on the relevant aberration. An image of the aberration should appear (see figure below)

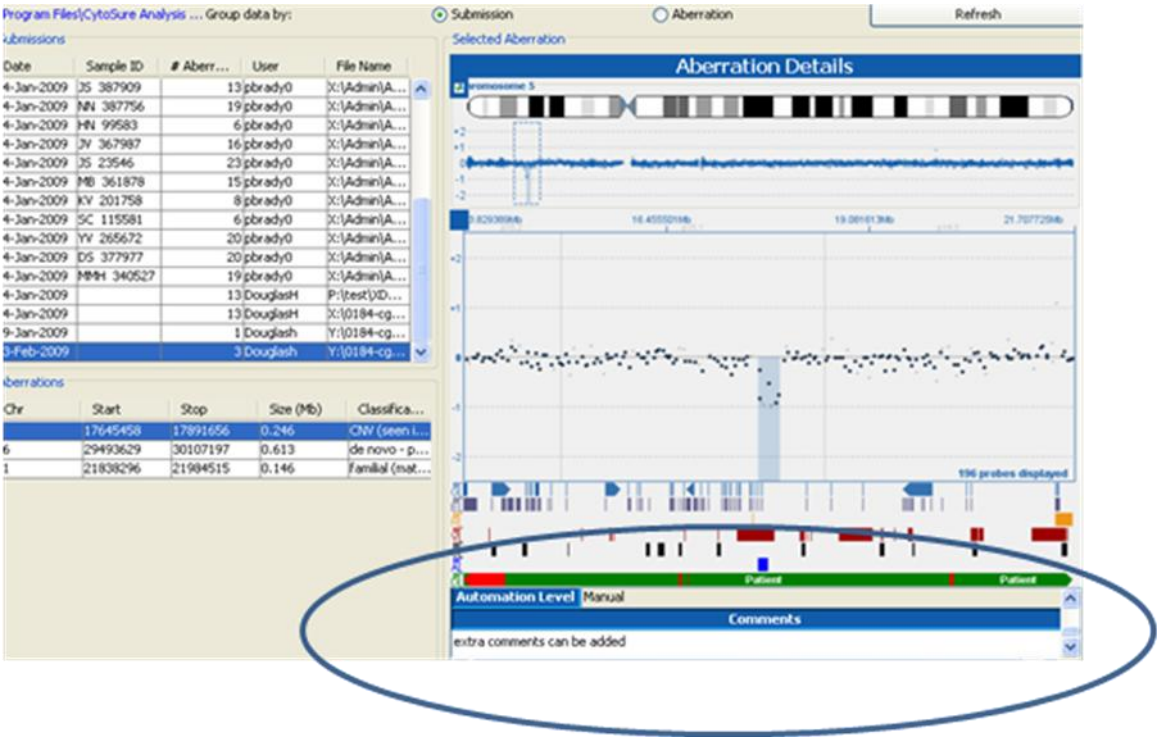
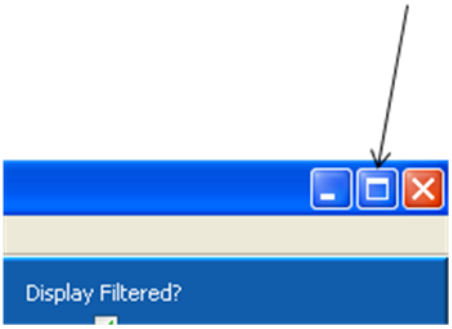


Figure 85 - Image showing the area where changes to the annotation can be made.

3. Below the image is a scroll bar where changes to the annotation (Classification and Comments) can be made. If the window cannot be seen it might be necessary to fully maximise CytoSure Analysis software window



4. Click ‘Apply changes’ to save the changes to the database.

5. In order to edit the aberration, click the commit button at the top right of screen; see figure 60).

7.5 Sorting or Filtration of data in the internal database

It is possible in CytoSure analysis software to filter aberrations that are stored within the database. For example a user might want to only see aberrations of a certain size within the database.

Sorting the data will simply order the data - for example it will order the data in size order with small aberrations first and large aberrations last.

The procedure is as follows:

1. In the database management screen, click on the data that is to be filtered or sorted.

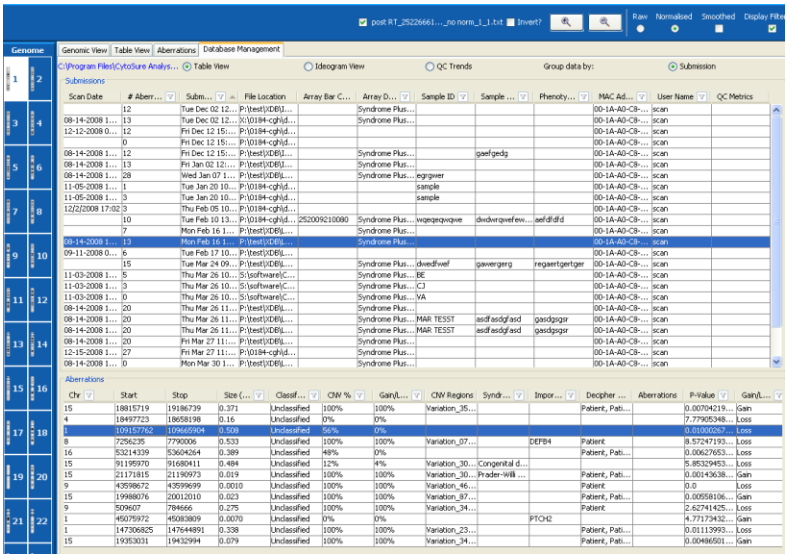


Figure 86 – Highlight the submission of the data to be sorted.

2. To filter the data. Next to the column header that is to be filtered there is a size icon

.

Click on this icon and a dialog box will appear

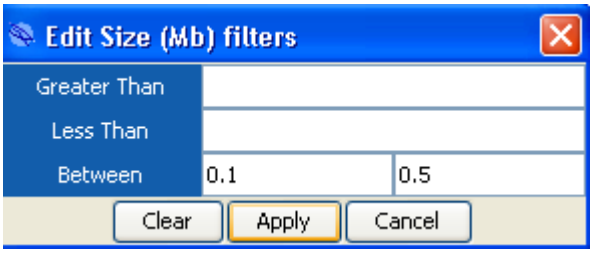
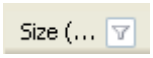
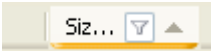


Figure 87 – Filter options.

Type in the parameters of the data that should be shown. For example, if the user wants to visualise only aberrations of a size between 0.1 and 0.5Mb, type 0.1 and 0.5 in the Between boxes, and click 'Apply'.

To remove the filtration, click on the 

icon - when the dialog box appears, press clear.

To sort the data, click on the column title. A new icon will appear 

alongside. Click on the to sort the data in ascending order. The icon will change to a and the data can be sorted in descending order.

7.6 Displaying QC metrics

QC metrics of data stored in the database can now be tracked over time. Navigate to the database modification page. Highlight the relevant submissions by holding the shift key and clicking on the submissions (see figure 88 below). Click on the QC Trends button (circled in figure 88 below).

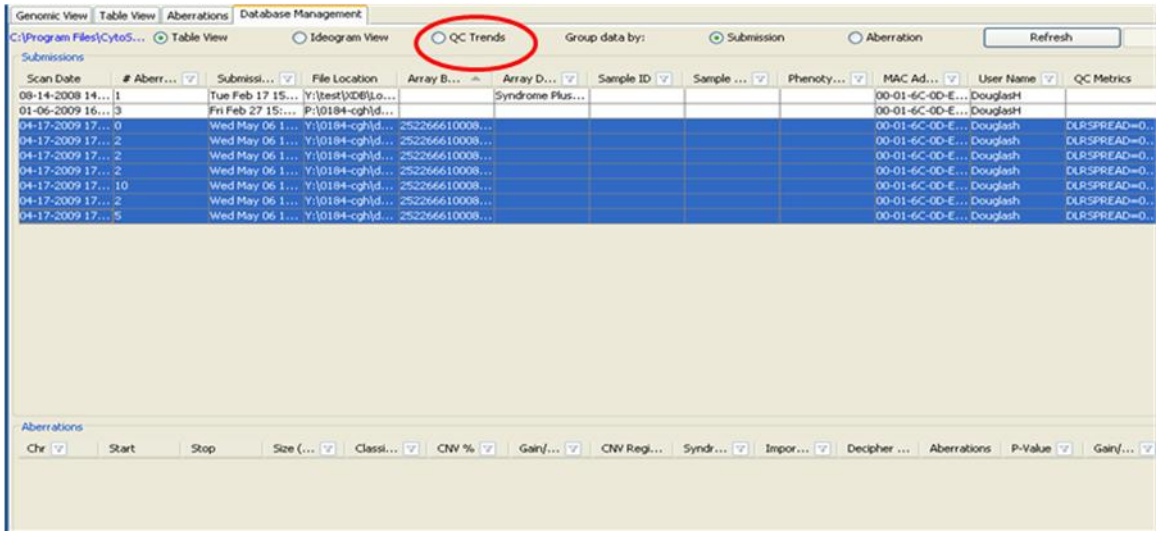


Figure 88 - Highlighting the relevant datasets.

QC metrics will be plotted on a graph. The X axis can be plotted in sequential order or by date.

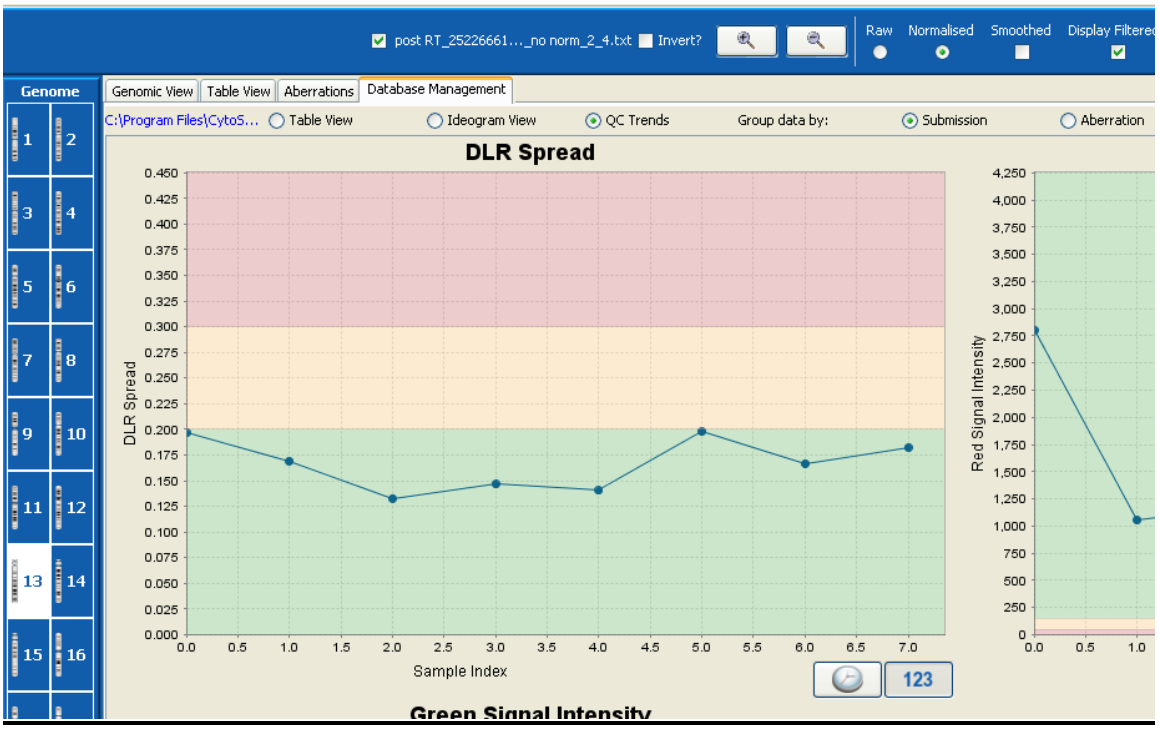


Figure 89 - QC metric plots.

7.7 Backing up the database file

To backup the database, click on the Backup button (see diagram).A backup database file will then be created to ensure that there is a copy in case the original db.xml becomes corrupted. In order to position the backup file, use Tools>Options> Files> Database Backup Location. If a previous version of the database is required, click Restore and a box will appear listing previously-backed up database files. Double-click on one of the files listed to carry out the restoration.

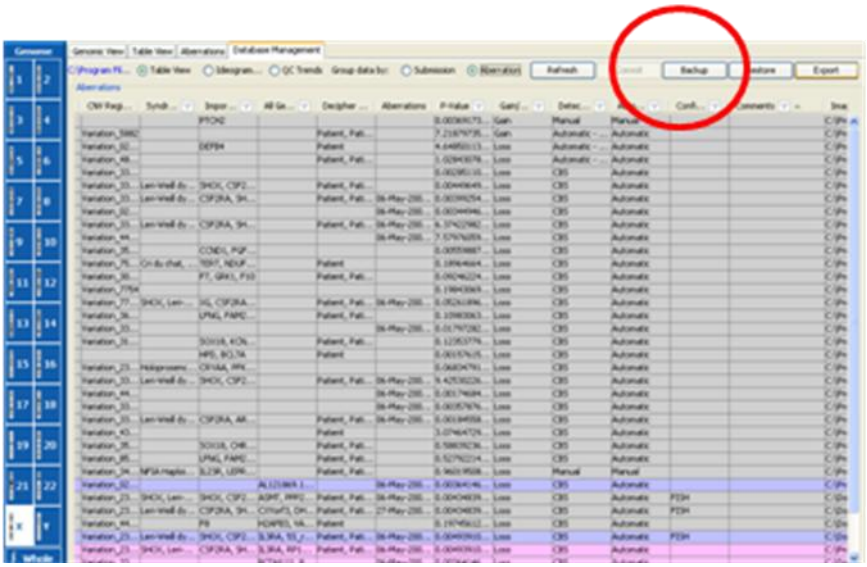


Figure 90 - Backing up the database file.

7.8 Exporting data in the database

There is now the ability to export the data from the database to a tab-delimeted .txt file. Clicking on the Export button in the Database Management tab will export the data. The .txt file can be opened using, for example, Microsoft Excel.

8 Population analysis

8.1 Grouping data and plotting aberration frequencies

The Population Analysis tab provides calculation and graphing of aberration frequencies from groups derived from your data. For example, it can enable you to query the database for the following data:

a) Which aberrations are present in my population with a different frequency in autistic samples compared to non-autistic samples?

b) Does this frequency vary in one ethnic group compared to an alternative ethnic group?

Question 1:

For population analysis it is important to type in the phenotype consistently in the phenotype box. So this can be done by either:

Import the .txt (or gpr) files into the software

Using File>Properties>Sample details

Editing the Phenotype Details field in the database

Figure 91 shows a view of a database with a number of samples containing the word “autism” in their Phenotype Details field. Note that there can be several terms included in this field (e.g. autism, developmental delay etc).

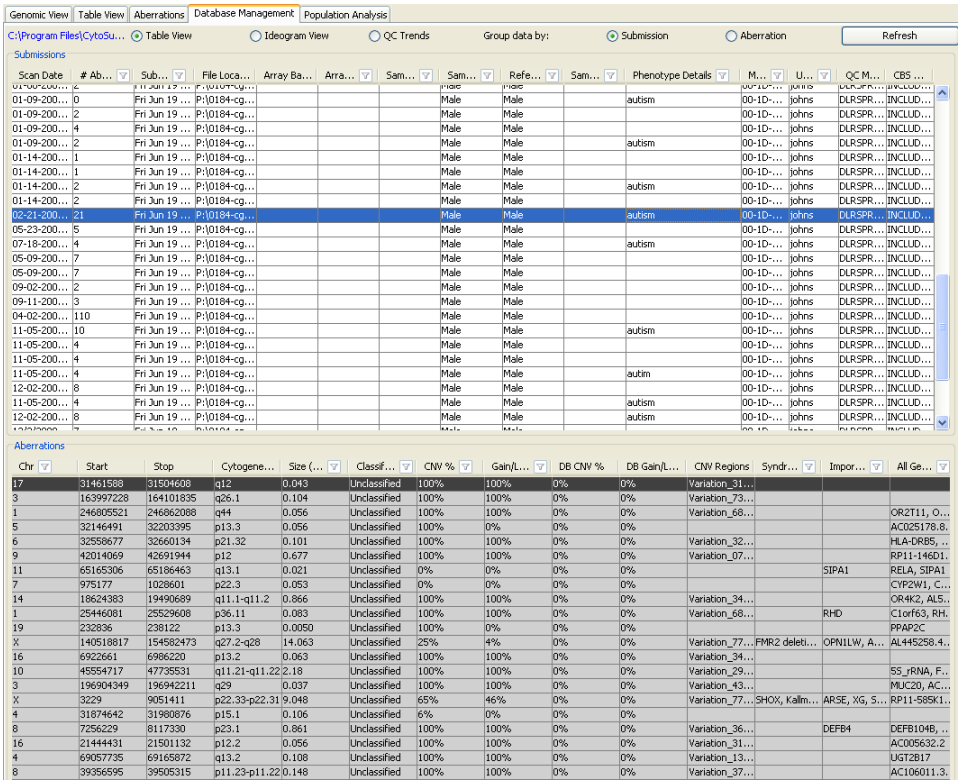


Figure 91 - Adding sample details to the database.

Once the details are added for all the relevant samples, click on the Population Analysis tab.

The first step is to subdivide the data into autistic samples and non-autistic samples. The screen is shown below

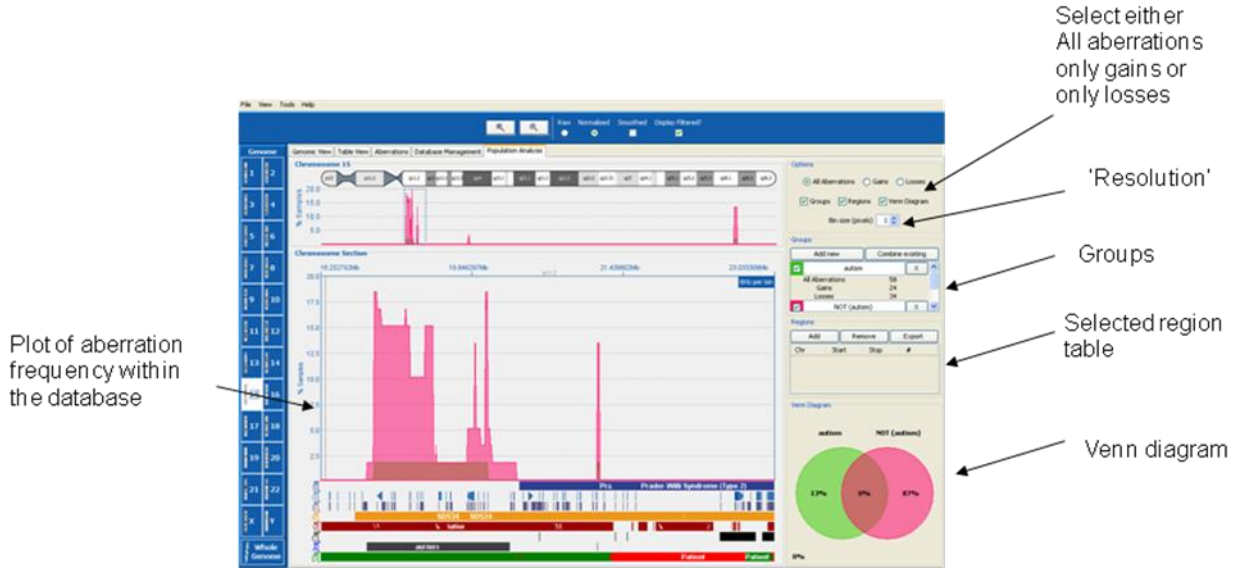


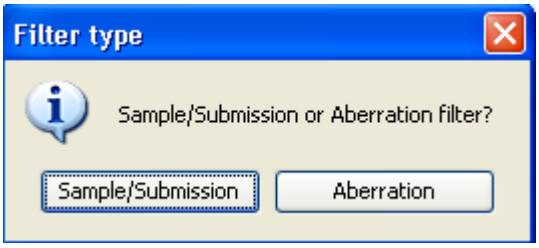
Figure 92 - View of population analysis page.

This shows all the data in the database plotted as aberration frequency (in percentage) versus genomic location. The data can be visualised as in the genomic view, so the chromosome buttons can be used to navigate through the chromosomes and the annotation tracks are functional so the position of genes can be identified.

The ‘resolution’ can be altered by increasing or reducing the pixels per bin.

In this first example, looking at the aberration frequency in autistic samples compared to non-autistic samples, the following procedure is used. The first step is to group the data into autistic samples and non-autistic samples.

1. Click on Add New in the Groups section



2. In this case the objective is to separate out the samples which have been labelled with “autism” in the Phenotype Details field from the remaining samples

Select Sample/Submission, select Phenotype Details from the drop-down box, type “autism” in the ”Contains” box and press Apply.

3. Allocate a unique name to this group. For this example “Autism” is appropriate.

4. A box appears prompting "Would you like to add the opposite group”. The opposite group will be all those samples whose Phenotype details do not contain the word “autism”. Therefore, these could be the normal population.

5. Click ‘Yes’ to add the opposite group (non-autism)

6. Allocate a unique name to this group.

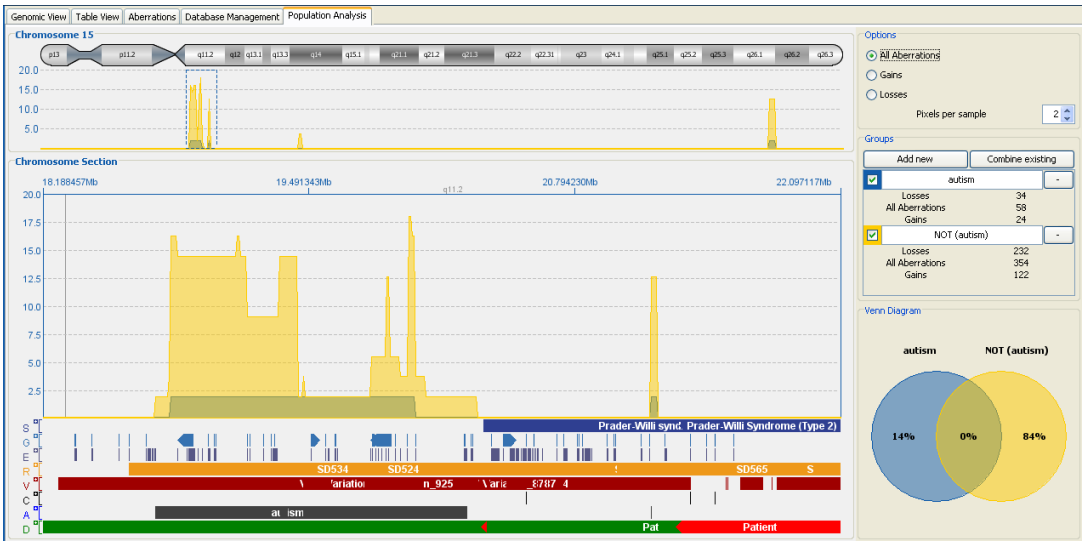


Figure 93 - Plot of aberration frequency.

8.2 Combining groups

Question 2: Combining groups

For this question the same example dataset is used, as above (autism and not autism). The software has the ability to determine the aberration frequency in autistic samples using different populations. See illustration below.

In this example, a population identifier (Latvia) has been added to the sample details

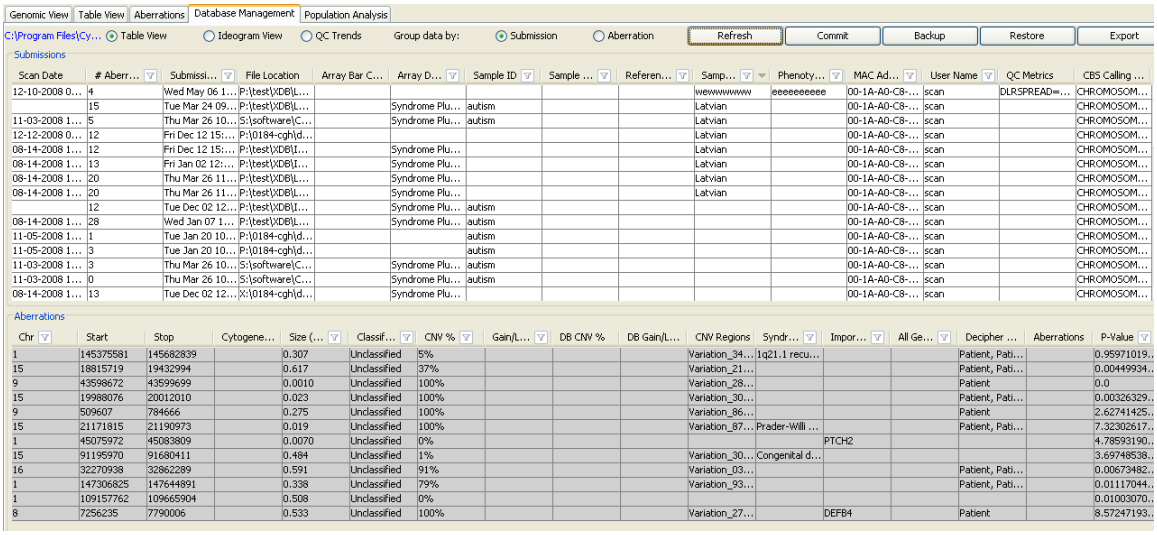


Figure 94 - Adding extra sample details to the database.

Now the samples are classified as Autism or not (Phenotype Details) and Latvia or not (Sample Details)

Following on from the previous example:

Click ‘Add New Group’
Select ‘Sample details’
Type in Latvia and click ‘Apply’

Name the group ‘Latvian’
Add the opposite group e.g. ‘NOT Latvian’

There are now 4 groups:

Autism
No Autism
Latvian
No Latvian

Note: The Venn diagram will show a maximum of 3 groups (the first 3 checked).

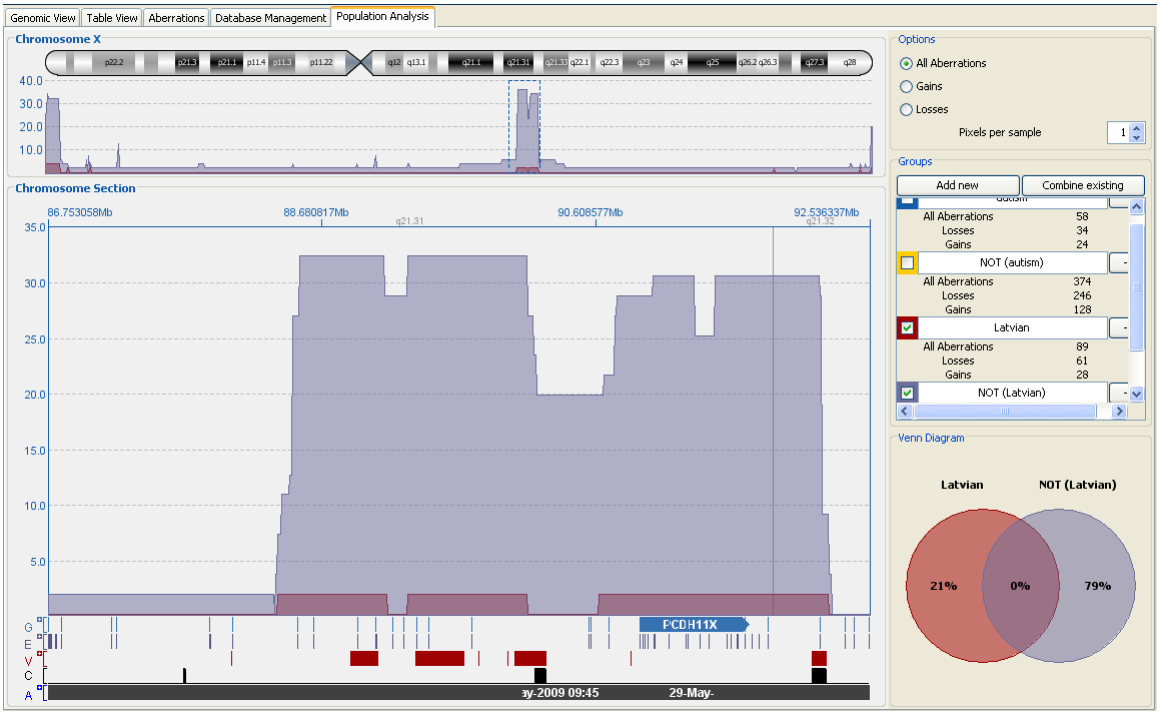


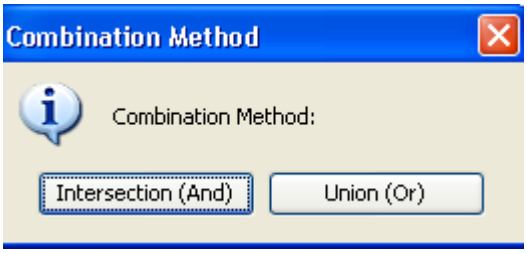
Figure 95 - Plotting extra groups

It is also possible to combine groups. To examine the aberration frequency in autistic samples in the Latvian population compared to the aberration frequency of autism in the non-Latvian population, complete the following:

Click ‘Combine Existing’
For Group 1 select ‘Autism’
For Group 2 select ‘Latvian’

For this example, the Combination method will be:

Intersection (And)



Click ‘No’ to opposite groups and ‘Yes’ to keep existing groups

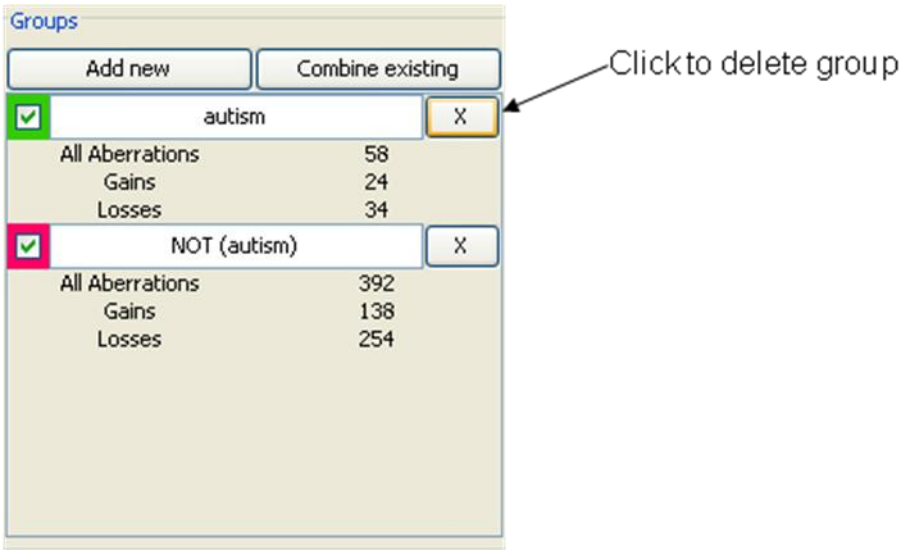
A group created will be: Autism and Latvian

A similar procedure of combining existing groups can be used to create a group of Autism and Not-Latvian.

The aberration frequencies are then be displayed on the graph.

8.3 Deleting groups

Next to the group name is a “delete” icon which can be clicked to remove a group.



8.4 Exporting aberration frequency results and statistics

To export the aberration frequency results, complete the following:

As in the Genome view, navigate so that the region of interest is present in the viewer

Click ‘Add’ to add that region to the list

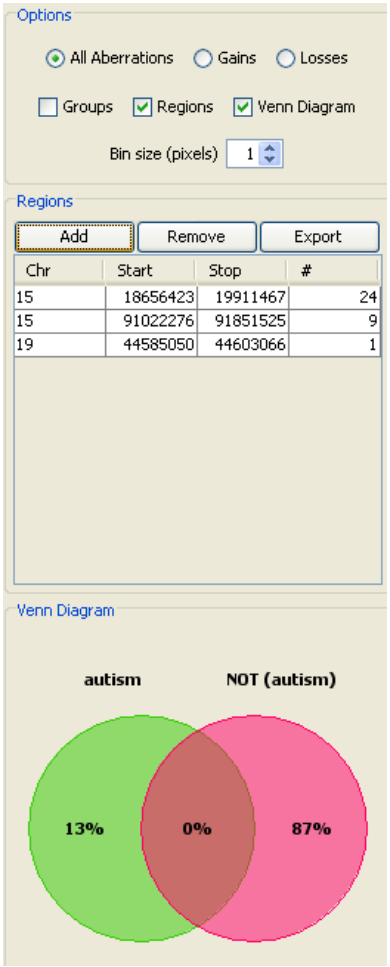


Figure 96 - Adding regions to the table.

To delete a region from the list, highlight the region to delete and click ‘Remove’.

The software incorporates some simple statistics for determining the confidence value of an aberration frequency result. There are two statistical tests, both of which will test if there is a significant difference between two groups. The tests are:

Chi-squared test
Fisher exact test (2 tailed).

A Fisher exact test will be used if in the contingency table an expected value is below 10, otherwise a Chi squared test will be used.

Click ‘Export’ and a .txt file will be generated with the following metrics:

1. Chromosomal position of selected genomic region.
2. Start position of selected genomic region
3. Stop position of selected genomic region.
4. Number of aberrations in the selected genomic region
5. Number of losses in the selected genomic region
6. Number of gains in the selected genomic region
7. Mean aberration frequency in the selected genomic region
8. Mean loss frequency in the selected genomic region
9. Mean gain frequency in the selected genomic region
10. Identity of the second selected group
11. P value for aberration frequency in the selected genomic region compared to a second selected group.
11. P value for a loss frequency in the selected genomic region compared to a second selected group
12. P value for a gain frequency in the selected genomic region compared to a second selected group

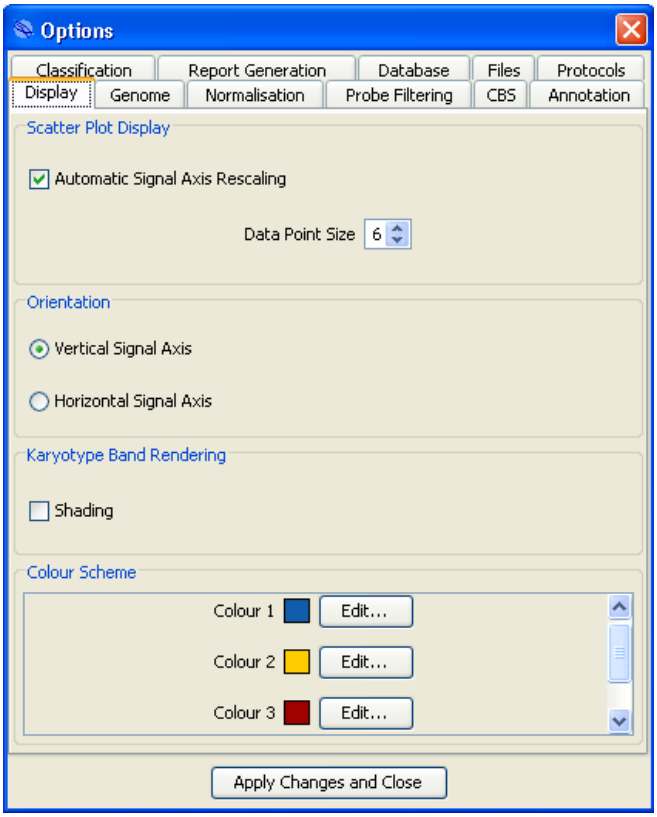
9. Setting up a protocol in Workflow mode

CytoSure now has workflow functionality. This means that all the various settings for carrying out an analysis are saved in advance as a protocol. It is anticipated that a fully trained ‘expert user’ would set up these settings to form a protocol

Once the protocols are created routine users can then follow a predefined workflow using the appropriate settings. A description about how to use the software in workflow mode is given in Section 10.

All the settings for setting up a protocol are located at Tools > Options. There are various tabs, a description of the options is given here.

Display



Automatic Signal Axis Rescaling: This setting will toggle on or off the y axis rescaling on the graph in the genomic view

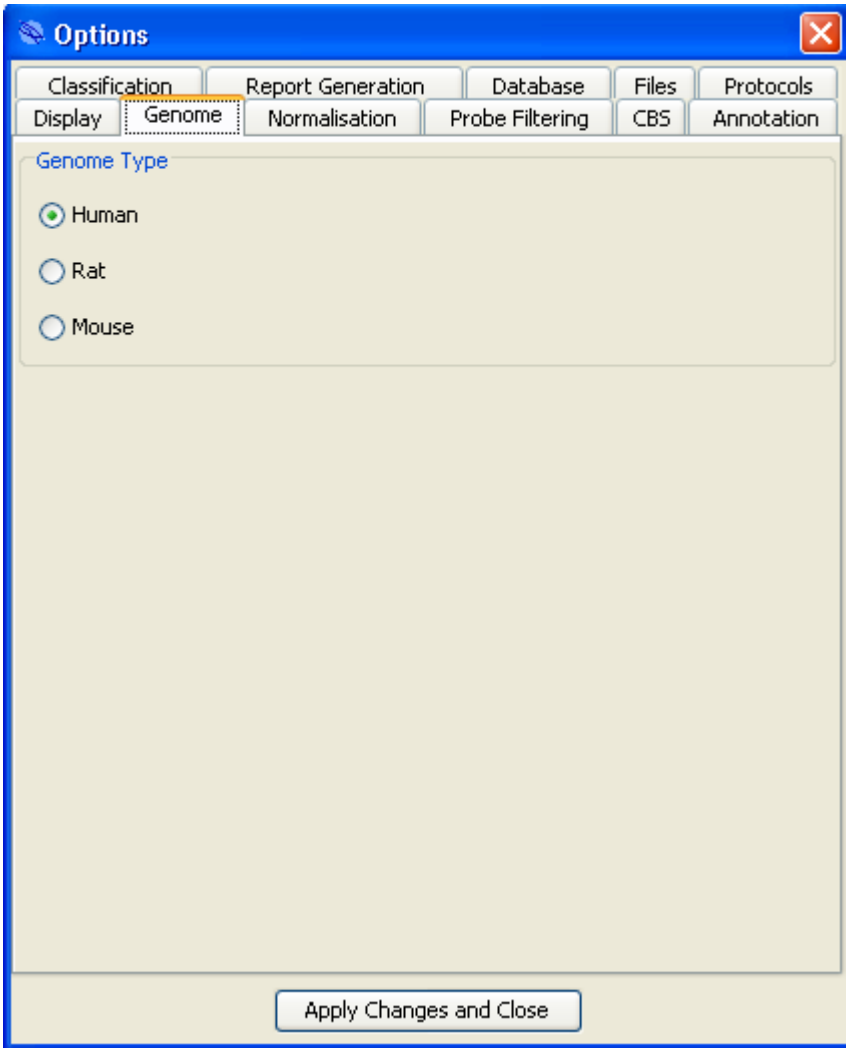
Data Point Size: Altering the setting changes the size of the datapoint in the genomic view

Orientation: This user can select the orientation of the graph. It can either be in horizontal or vertical view (see section 4.8)

Karyotype Band Rendering: This sets the shading in the chromosome diagram on the ideogram

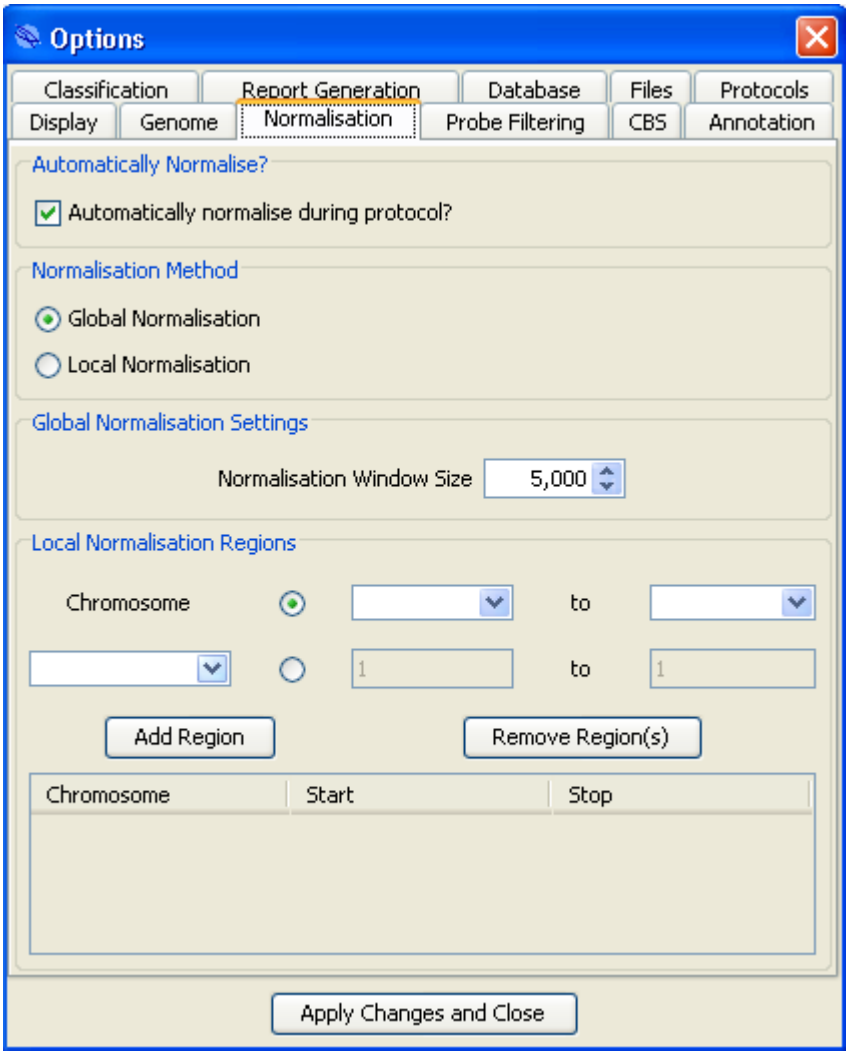
Colour Scheme: Allows the user to set the colour of the datapoints

Genome Setting



The user can set the type of annotation used

Normalisation setting



Automatically normalise during protocol:

Some feature extraction software will normalise the data. If the data has been normalised by the feature extraction software then the tick box should be ticked off. If the data has not been normalised then the tick box should be ticked and the CytoSure software will normalise the data

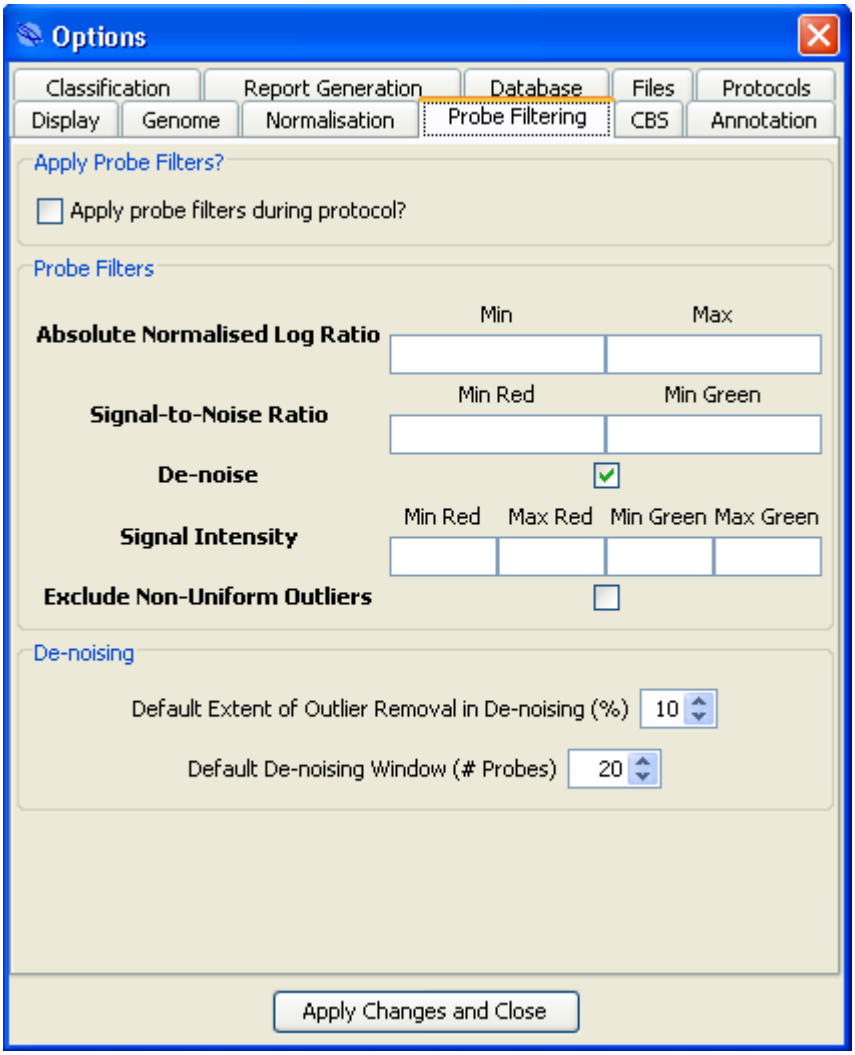
Normalisation Method: Assuming that the CytoSure software is carrying out the Normalisation then for most arrays the Global Normalisation should be selected. However, with some custom arrays, where there is the possibility that an aberration could cover are large proportion of the probes on the array, then the local normalisation should be selected.

Global Normalisation Setting: This setting should not need to be changed

Local normalisation method: If local normalisation is selected then the region of the genome that is used for normalisation needs to be selected. This should be a region that is expected to have a log2 ratio = 0. The chromosome should be selected and then the region can be selected by band or genomic position. Then click on 'Add region'

Probe filtration setting

Probe filtration is the ability to remove certain probes from the analysis because they may have low signal or certain other non-desirable characteristics

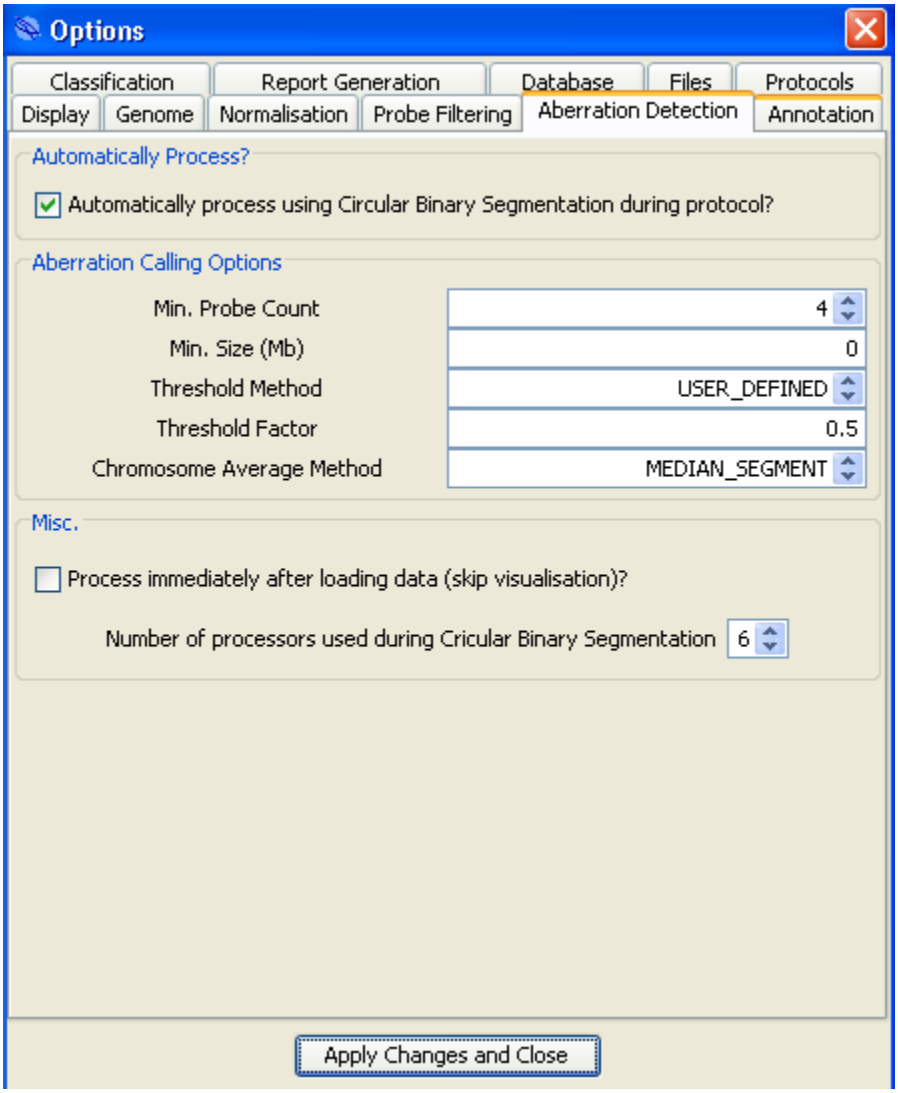


Apply Probe Filters? Should the display and the data be filtered (see section 4.5)

Probe filters: Select type of probe filtering (see section 4.5 for more details on probe filtering)

Denoising: If denoising is selected then the extent of the denoising can be selected here. Denoising removes a certain % of probes at either end of the log2 ratio spread in a defined window. These metrics can be altered here.

CBS settings



The CBS settings can be altered here.

Minimum Probe Count:
A value of 4 probes is suggested as the minimum number of probes in a segment that are required to make a call.

Threshold method and threshold factor:
This is the value that the algorithm uses to set the threshold value.

User definable – Type the value of the $\log_2$ ratio into the threshold factor box to set the threshold value that will be used to make a call. Typically a $\log_2$ ratio of 0.5 to 0.85 is used.

Standard deviation – the software calculates the standard deviation of the ratio ; the threshold value will be the threshold factor x standard deviation of the ratios. The preferred factor is 3 or 4.

Deviation Log Ratio (DLR) – the software calculates the DLR Spread of the ratio ;the threshold value will be the threshold factor x DLR of the ratios. This typically is a factor of 3 or 4.

X separation – The software uses the value from the X chromosome probes. This assumes that a sex mismatch experiment has been run. It may omit duplications, which usually need a lower threshold value compared to deletions.

Chromosome average method:
The method of averaging the segments at the baseline. This is used to set the threshold value.

Median segment uses the Median value of all segments in the chromosome as the baseline, and is the preferred option.

Mean segment uses the Mean value of all segments in the chromosome as the baseline. This could be affected by large aberrant regions in the chromosome.

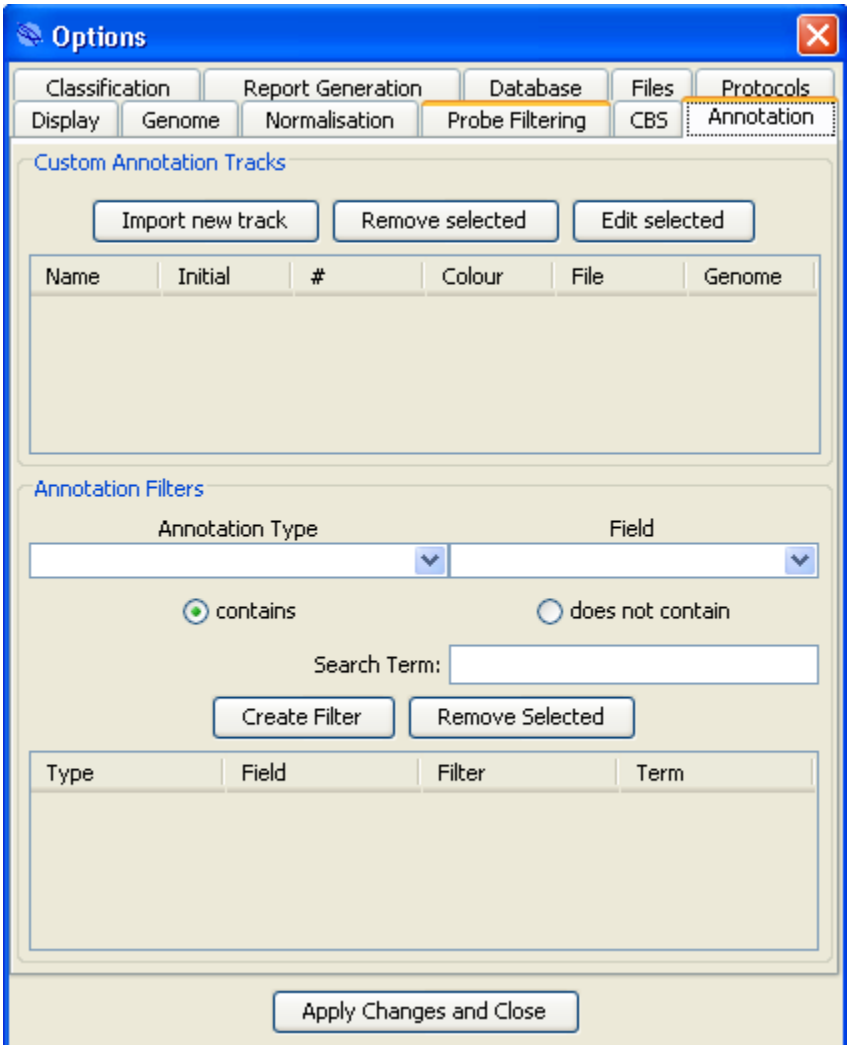
Zero will set the baseline at zero whatever the data is.

Chromosome Mean log probes – Uses the mean value of all the probes as the baseline

Chromosome Median log probes – Uses the median value of all probes as the baseline

Annotation settings

In this section the user can alter the information displayed in the annotation tracks



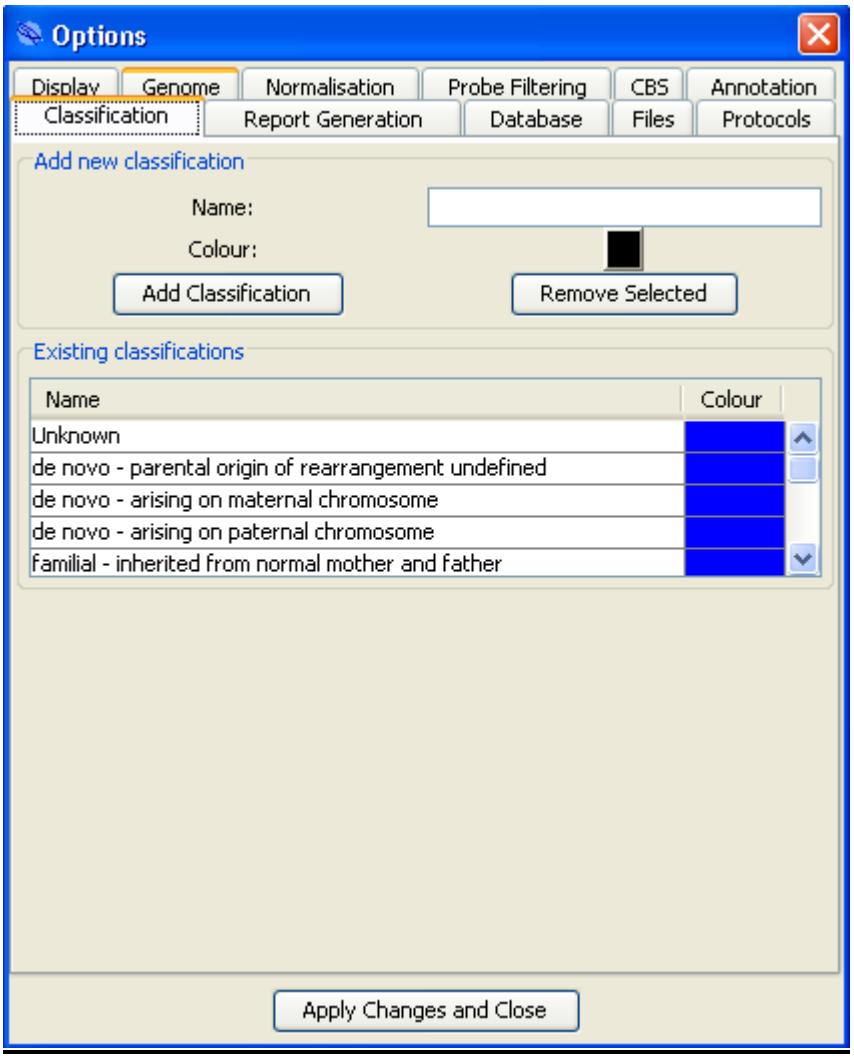
Custom annotation track

The user can add a custom annotation track. The user will have to create a .txt file with a header containing Chr – Start – Stop – ID. Then click 'Import New track' and navigate to the relevant file. More details about how to set up custom annotation tracks are described in section 4.2.

Annotation Filters

The annotation can be filtered to remove certain entries. For example from the CNV DGV annotation the user may wish to remove CNVs that are inversions. Detailed information about annotation track filtering is given in section 4.2

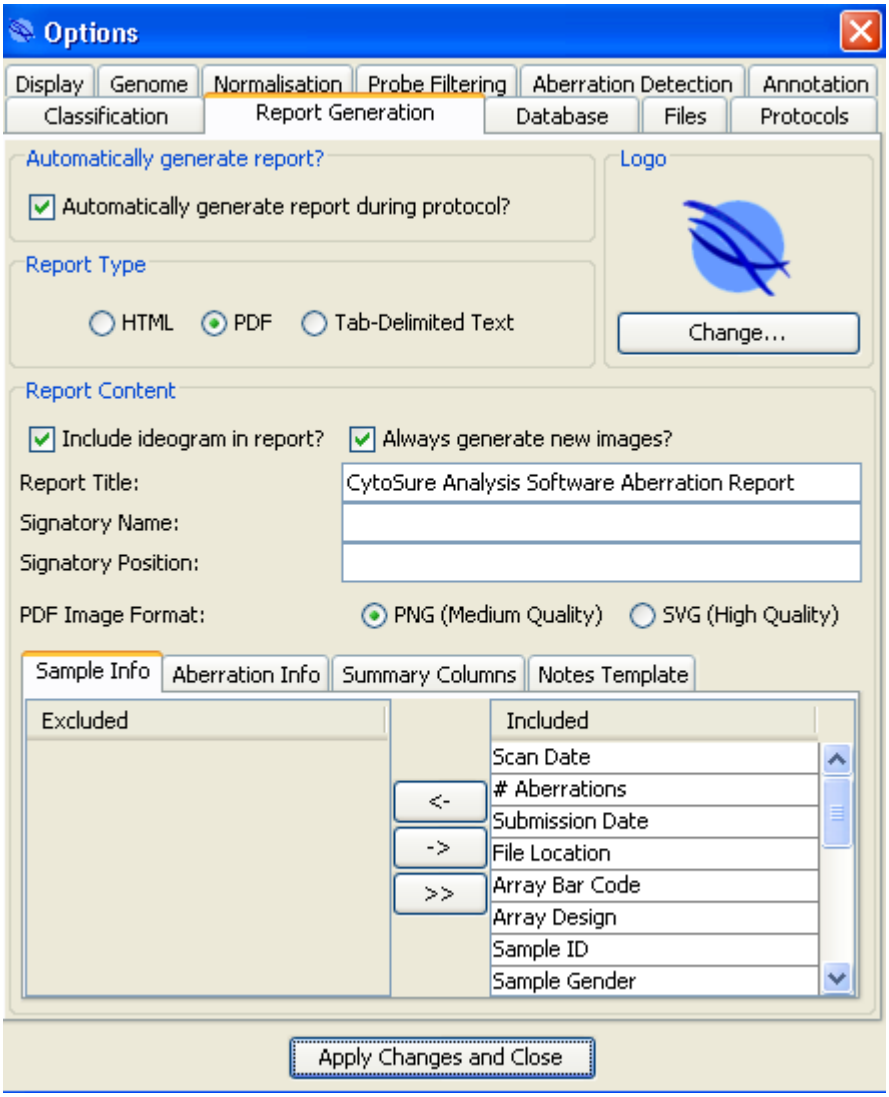
Classification settings



In the aberration table (see section 5.5.2) it is possible to classify the aberration such as CNV, de novo etc. In this section the user can define their own classification terms and what colour the annotation blocks should be.

Reporting settings

This section outlines the customerisation of the report.



Automatically generate report during protocol?

If ticked, then in workflow mode, the software will automatically generate a report after the user has annotated the aberratons.

Logo

A logo can be added to each report. Click on the ‘change’ button and navigate to the file containing the logo. The logo can be in a JPEG or BMP file.

Report type

The user can select the type of report format. The HTML format can be opened in Microsoft Word as is therefore editable.

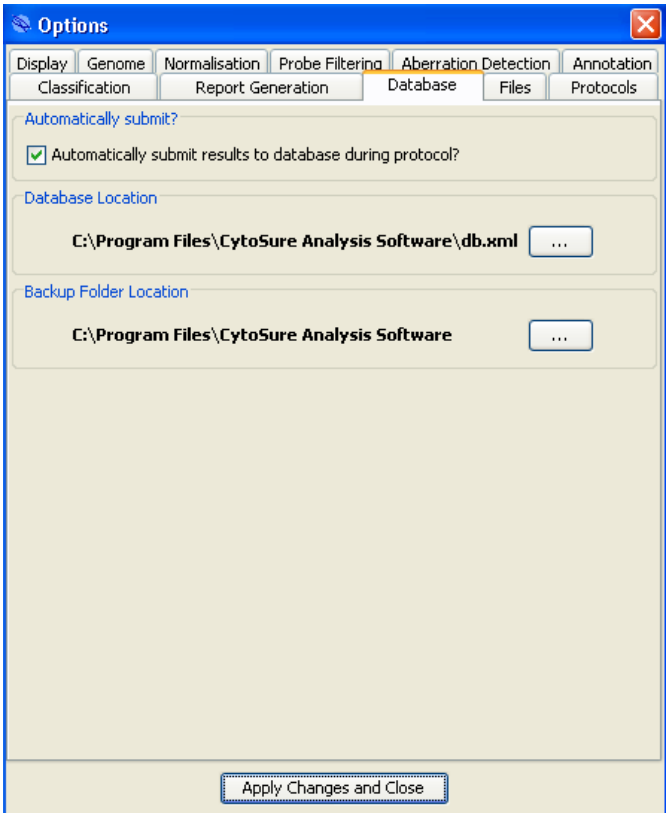
Report content

The Report title will place text at the beginning of the report

The Signatory name and position will place text at the end of the report

The user can define which information they would like to include in the report. If the Notes tab is selected then standard text can then be included in each report. More details about the report are discussed in section 5.6

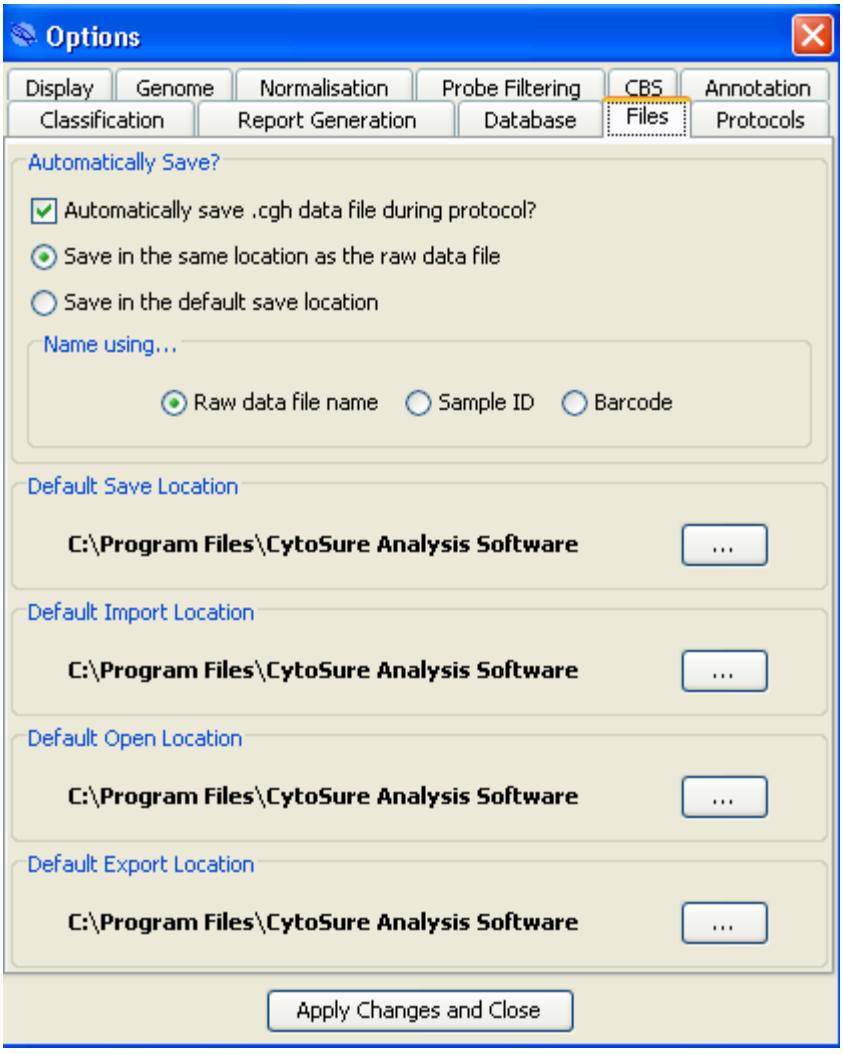
Database settings



The user can specify whether the aberration results are submitted to the database during the workflow mode.

The location of the database file and the backup file can also be specified.

Files settings



Automatically save

If ticked, the software will automatically save a .cgh file after the reporting. There are also options here to set the location where the .cgh file will be saved.

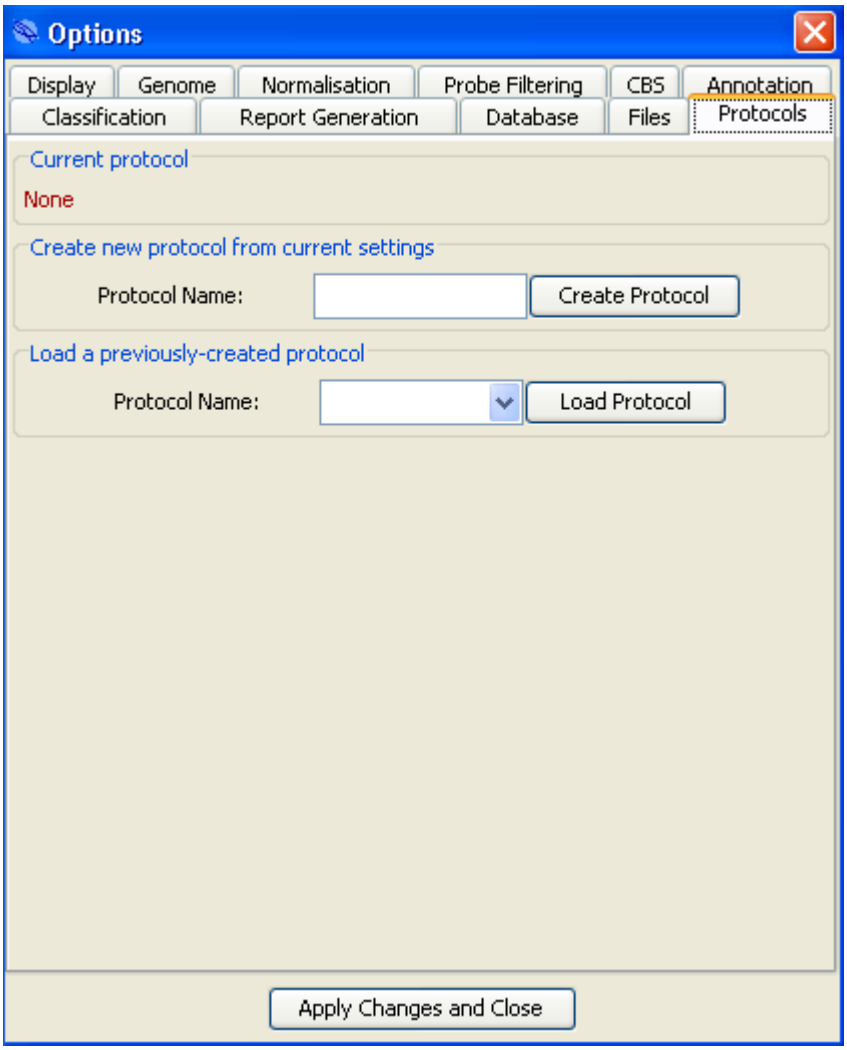
Name using

There is a choice of naming the file as either the same file name as the .txt or .GPR file that has been used. Alternative naming is by the sample ID or by the barcode of the slide.

The default file locations

This specifies the folders where the files to be loaded into the software or alternatively where the files exported from the software should be located.

Protocol settings



Once the user has set all the various settings these settings can be saved as a protocol. Type in a protocol name and then click ‘Create Protocol’.

The user can then Load in a previous protocol by selecting the relevant protocol in the drop down menu to restore all the previous settings.

If during the analysis the user changes any of the settings, then the Current Protocol will change to a ‘None’

11. Use of the software in Workflow Mode

Routine users can use the software with the settings set as a protocol.

The principle of this process is that the analysis should be easy to use and can be standardised between multiple users. All the settings will be the same. However, it is still possible for users to change the settings, however there will be 2 consequences:

1) There will be a warning asking the user if they want to exit the workflow mode
2) In the database and the report the analysis protocol will not be stated.

There are the following steps when running the software in Workflow mode

1) Select protocol
2) Load
3) Process
4) Describe
5) Review
6) Classify
7) Report
8) End

The buttons are across the top of the screen allow the user to move from one step to another

1) Select protocol

Various protocols should have been set up using the options highlighted in section 10. Included is the default protocol which contains the OGT settings. Click on the Protocol button at the top left of the screen and select a protocol. It is important to press ‘Apply Changes and Close’ button.

2) Load

Click on the load button and an option comes up to select either feature extracted files or batch processed files.

Feature extracted files

These should be feature extracted files and so will either be .txt (in the case of those generated by Agilent feature extraction software) or .gpr files (in the case of GenePix). Multiple datasets can be loaded. The number depends on the computer and the type of array. Typically 3 – 4 datasets of a 105k array can be loaded. Once loaded the dataset will be normalised (if appropriate tick box in the Aberration detection is ticked in the protocol). The dataset will then be visible for an initial look at the data (again if the appropriate tick box in the Aberration detection is ticked in the protocol)

Batch processed files

Feature extracted files can be CBS processed in a batch. To batch process feature extraction files, click Tools -> Batch processing. Then click on the 'Change' button and select the protocol. Following the batch processing, the files can then be loaded into the workflow using the 'Load into workflow' button. Alternatively the processed files can be accessed directly in the workflow using the Load button.

Display of files in the workflow

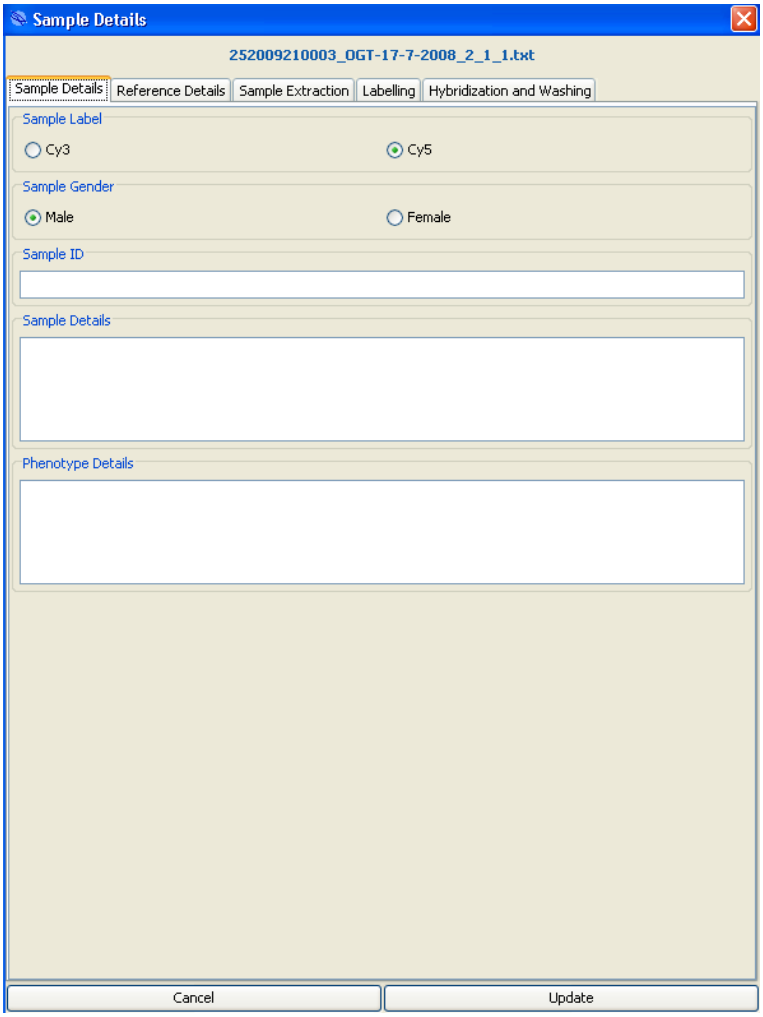
Files that are present in the workflow mode are marked with a star

3) Process

On clicking the process button, the data will undergo CBS processing, and then the aberrations will be called using a threshold. This step will be faster on a quad core computer (see system requirements) and should take about 3 - 4 minutes for processing a 2x105k Syndrome Plus array.

4) Describe

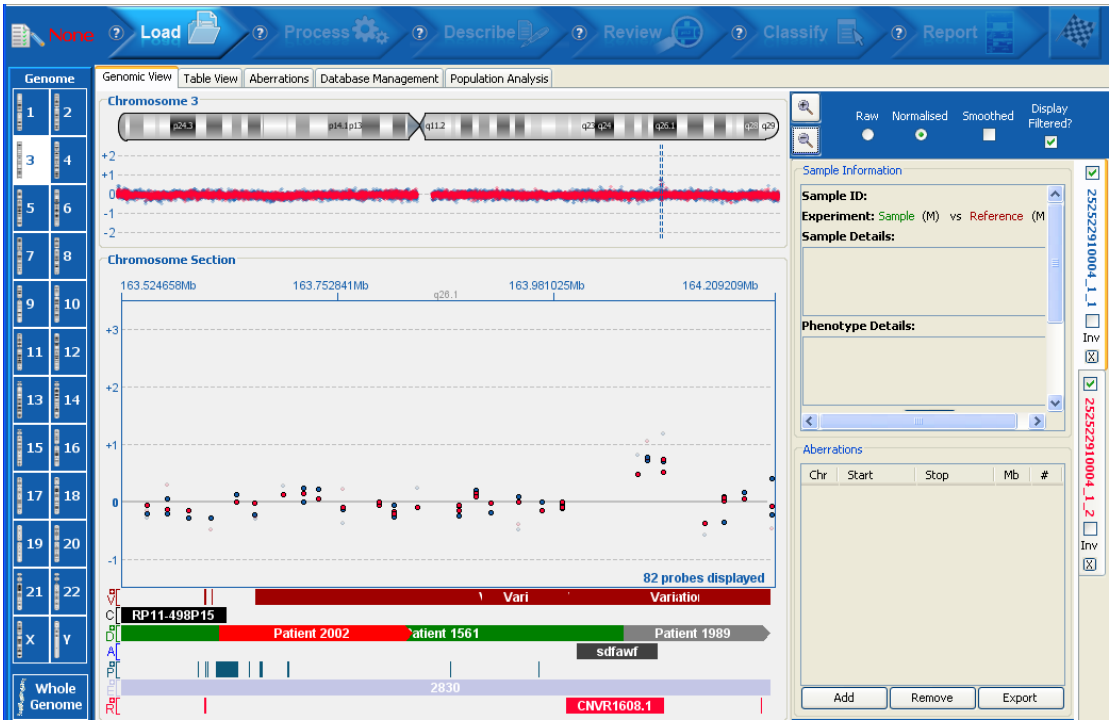
During this step the user can type in the sample Id, the sample description and sample phenotype. Additional information can be added by clicking on the tabs. If more than one dataset is loaded then the filename will be highlighted with the relevant colour and once update is pressed a second box will appear. When complete click ‘Update’



5) Review

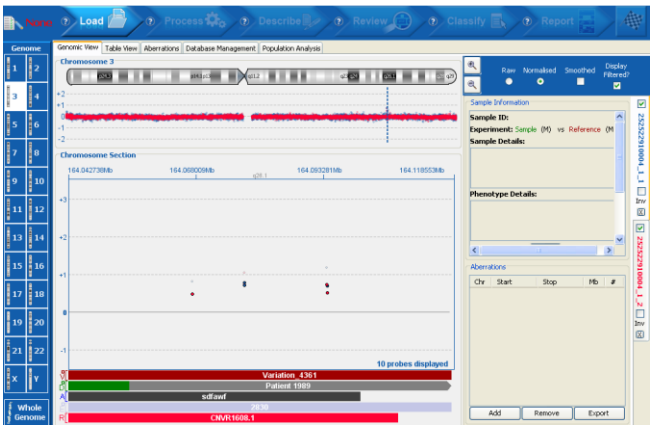
The aberrations that the software has detected, can then be reviewed by the user. Manual aberration detection can be used to either add additional aberrations or delete them. This process is described below.

a) Locate the aberration in either the chromosome view or whole genome view.



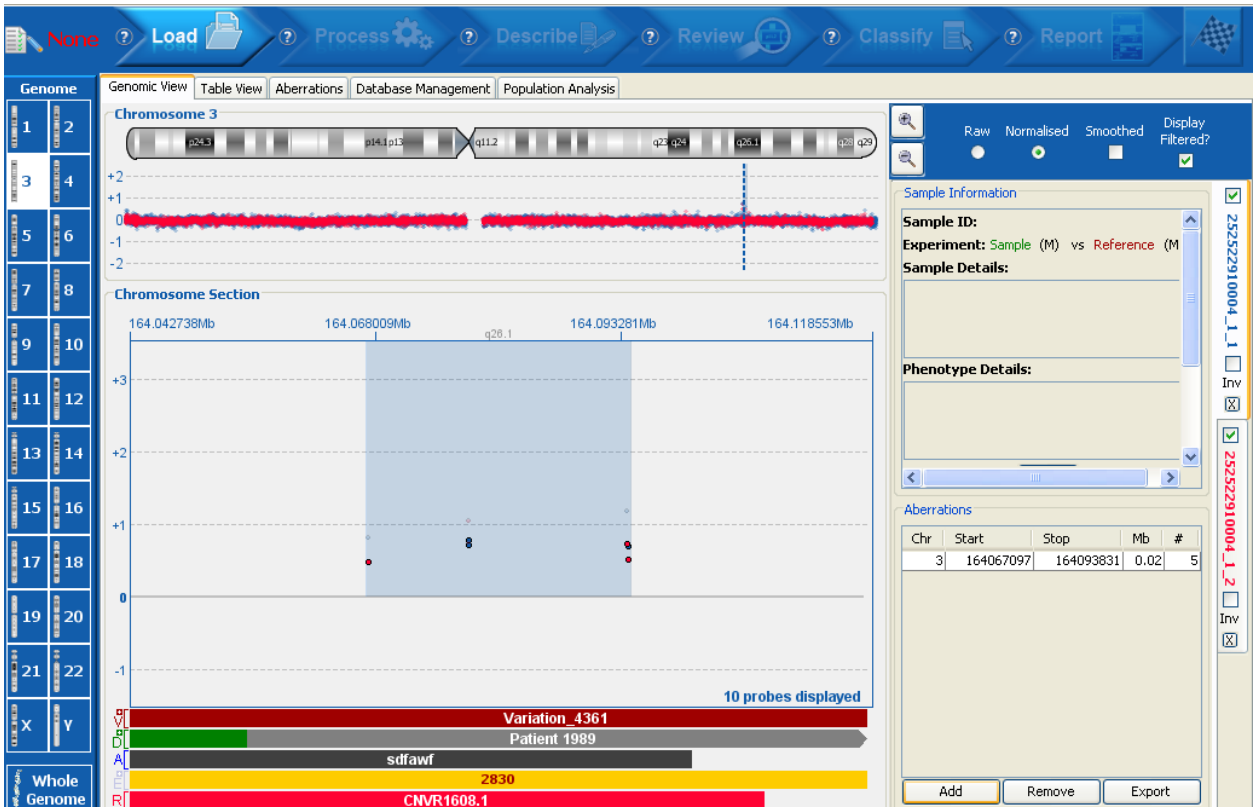
Locate the aberration.

b) Zoom in. **All probes shown on the screen should be within the aberration**.
There should be no flanking probes visible; otherwise these will be included in the export.



Expand so only the data points in the aberration are visible on the screen.

c.) Click ‘Add’. Repeat for other aberrations in the chromosome. **It is important to include all aberrations on the same chromosome (see discussion below)**



Classify

Click on the classify button to progress to the next stage. This is the process where the user needs to classify the aberration according to whether the aberration is a CNV (ie not involved in disease) or a potential pathogenic aberration that requires further investigation

The software has a suite of tools to help. Principally the following questions need to be answered

1) Is the aberration covered by a CNV entry in the DGV? If it is then it is likely to be a CNV

2) Is the aberration where a Syndrome aberration is?. - If it is then it is likely to be pathogenic

3) Does Decipher or Ecaruca contain similar aberrations? – If so then the aberration could be pathogenic

4) Do any of the function of genes suggest they might be involved in disease? – If so then the aberration might be pathogenic

5) Does the aberration have 'important genes'? If it does then the aberration may be pathogenic

6) Have you seen the aberration before? If you have with similar phenotype, then the aberration might be pathogenic

The annotation tracks are in key tool in answering many of the questions. They are described in more detail in section 4.2

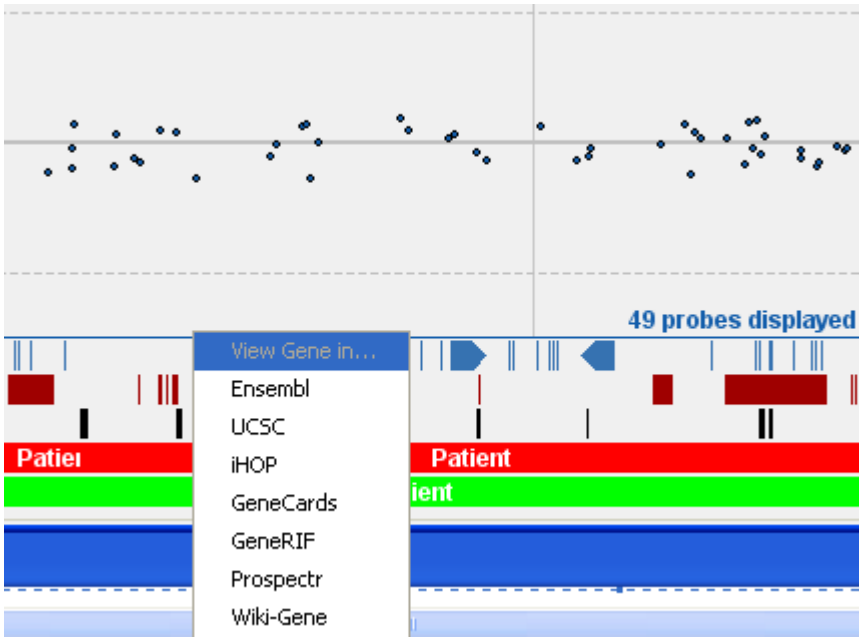
Question 1
To determine if the aberration might be a CNV then the annotation tracks are used. The detailed use of the annotation tracks is described in section 4.2. The DGV track (dark red) contains information about whether a particular aberration is present in the normal population. Therefore, if there is good coverage of an aberration in this track then the aberration should be covered as a CNV.

Question 2
The Syndrome track (dark blue track) shows the position of aberrations found in common cytogenetic syndromes. Clicking on the block will take the users to more information about the syndrome

Question 3
Decipher and Ecaruca are on line databases containing information about cytogenetic aberrations. There are Decipher and Ecaruca annotation track which show the position of these aberrations. The aberration might be coloured red if the aberration is an deletion, green if it is a gain and grey if this information is not available. Clicking on the Decipher blocks will take the user to the Decipher web site with more information.

Question 4
Using the right click on the shaded region of an aberration will take the user to the SUSPECTS site. This has several tools to help the user decide if any of the genes might be involved in disease

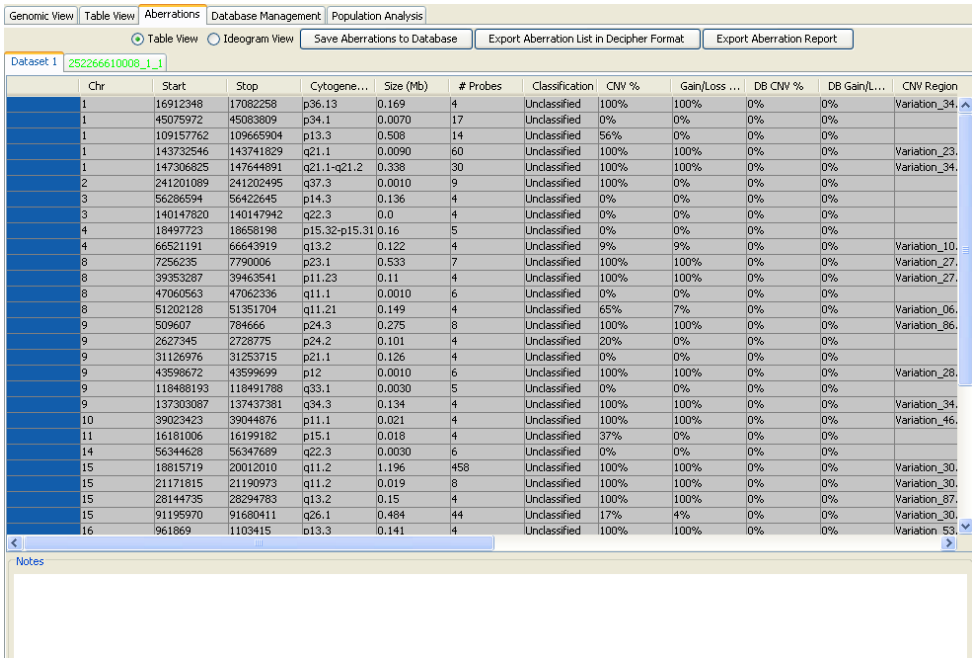
Right click on the gene. This will bring up some options of sites with publically data on the function of genes. See section 4.2 which gives more details on these.



Question 5
The software stores disease-causing genes from OMIM and will check the aberration for whether they are present in a particular aberration. To determine this, then click on the Aberration tab, which will open a table of the aberrations.

Question 6
As the user carries out more experiments, the software will save the aberrations and the classifications to a database. This data will then be displayed on an annotation track. This is track D. The aberrations will be coloured according to the classification. Therefore normally an aberration which was previously classified as a CNV will coloured a different colour from an aberration that was previously classified as potentially pathogenic.

Use of the population tab (see section 8) display the frequencies of the aberrations in the database. The data can then be subdivided on the basis of phenotype and/or population for detailed study.



One of these columns contains information about if the aberration contains any important genes.

Classification

With these tools it should be possible to classify each aberration. To do this click on aberration tab, then right click on the aberration and the various classifications will come up. Select the relevant aberration

Further information can be added to aberration table, such as in the Notes page, then text can be typed in

Report

Once the classification is complete, then clicking on the report button will carry out the following processes:

1. Save the aberrations to the database
2. Produce the report in the format specified by the protocol
3. Save the classification and annotation as a .cgh file. This can be opened using File>Open.

Both the report and the database will contain a record of which protocol was used to carry out the analysis.

Finish

Clicking on the Finish button will complete the process and the software will now be ready to load new datasets. The user can click on the Finish button and the dataset will be closed (following a warning).

Altering the protocol in the middle of workflow mode

Changing the protocol in the middle of workflow mode or altering the settings will result in the software producing a warning. The data can still be analysed using the revised settings, however, the report and the database will not contain a record of an analysis protocol used.

12. Other features

Array metrics
X chromosome separation is shown at the top right hand corner of the **chromosome section view** (C). It calculates the mean ratio of the Chromosome X probes. It is intended for experiments with male and female DNA when the Chromosome X probes are being used as a control.

Other array metrics such as mean signal intensity are available by clicking **File** -> **Properties.**

Exporting data and images
Data from CytoSure can be exported by using **File** -> **Export**. A choice is then offered between an image and a text file. The text file contains the normalised data.

File defaults

CytoSure uses default settings for importing and saving files. These default locations can be altered by clicking **Tools** -> **Options** and then clicking on the **Files** tab. Click on … and navigate to the folder required.

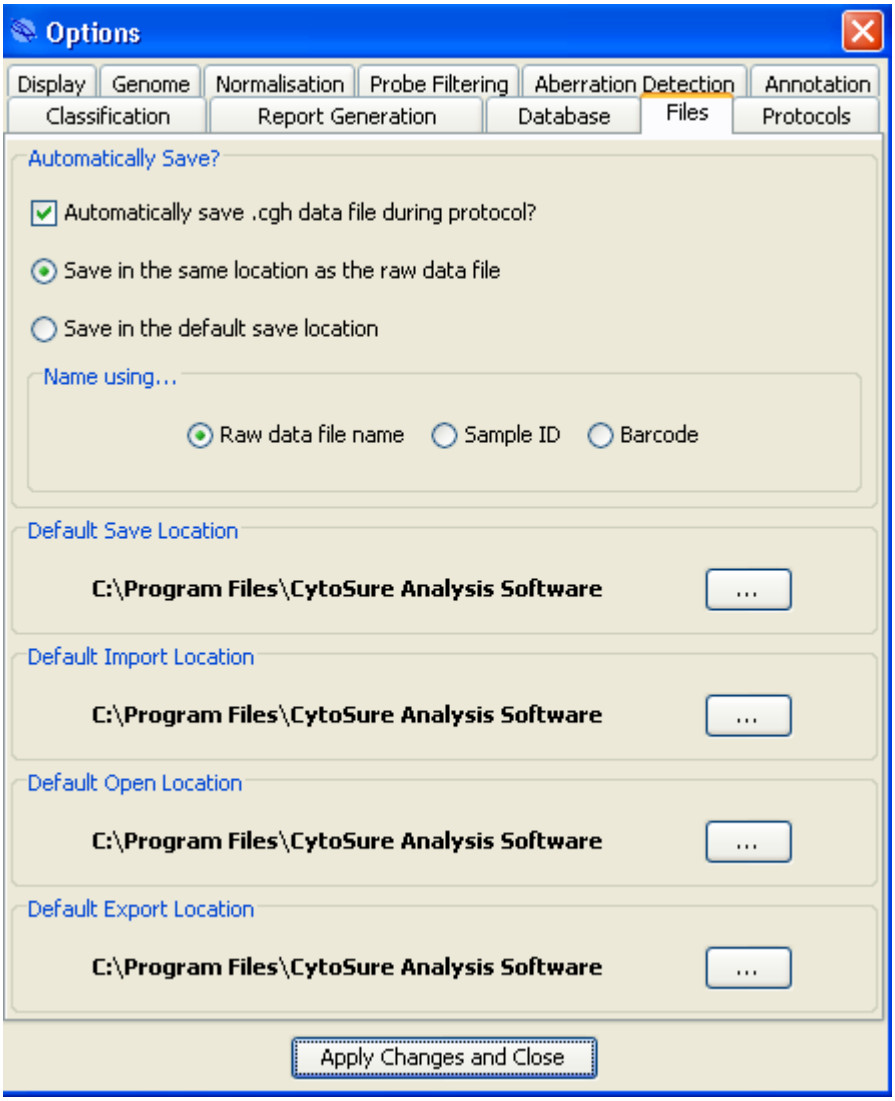


Figure 83 – Setting the file defaults.

13. Contact details:

Oxford Gene Technology, Begbroke Science Park, Sandy Lane, Yarnton, Oxford OX5 1PF, UK

Oxford Gene Technology (Operations) Ltd. Registered in England No: 03845432 Begbroke Science Park, Sandy Lane, Yarnton, Oxford OX5 1PF, UK.

T: +44(0)1865 856826 ; F: +44 (0) 1865 848684 ; E: products@ogt.co.uk; W: www.ogt.co.uk

Technical support

E: support@ogt.co.uk

14. Legal Information:

Trademarks: OGT™ CytoSure™